DOCUMENT RESUME

ED 425 194 TM 029 264

AUTHOR	Burke, Kay, Ed.
TITLE	Authentic Assessment: A Collection.
ISBN	ISBN-0-932935-51-6
PUB DATE	1992-00-00
NOTE	228p.
AVAILABLE FROM	IRI/SkyLight Training and Publishing, Inc., 2626 S. Clearbrook Drive, Arlington Heights, IL 60005; Tel: 800-348-4474 (Toll-Free); Fax: 847-290-6609; e-mail: info@iriskylight.com; Website: http://www.iriskylight.com
PUB TYPE	Books (010) -- Collected Works - General (020)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Budgeting; Curriculum Development; *Educational Assessment; Educational Trends; Elementary Secondary Education; *Performance Based Assessment; *Portfolio Assessment; Standardized Tests; *Student Evaluation; *Test Construction; Test Use
IDENTIFIERS	*Authentic Assessment

ABSTRACT

The essays in this collection consider the meaning of authentic assessment and the implications of its use. "Section 1: Assessing Assessment" includes the following essays: (1) "The Nation's Report Card Goes Home: Good News and Bad about Trends in Achievement" (Robert L. Linn and Stephen B. Dunbar); (2) "Budgets, Politics, and Testing" (Chris Pipho); and (3) "Why We Need Better Assessments" (Lorrie A. Shepard). "Section 2: Testing and Thoughtfulness" includes the following essays: (4) "Testing and Thoughtfulness" (Rexford Brown); (5) "Project Spectrum: An Innovative Assessment Alternative" (Mara Krechevsky); and (6) "Teaching to the (Authentic) Test" (Grant Wiggins). "Section 3: Alternatives to Testing" includes the following essays: (7) "Searching for Alternatives to Standardized Tests: Whys, Whats, and Whithers" (Walter Haney and George Madaus); (8) "Assessment Literacy" (Richard J. Stiggins); (9) "Performance Assessment: The Message from England" (Desmond L. Nuttall); (10) "But...the Curriculum" (Carolyn Mamchur); (11) "Linking Restructuring to Authentic Student Achievement" (Fred M. Newmann); (12) "What Teachers Learn from 'Kid Watching'" (Richard VanDeWeghe); and (13) "Assessment: A Thoughtful Process" (Steven Ferrara and Jay McTighe). "Section 4: Journals and Portfolios" includes the following essays: (14) "Assessing Reflective Responses in Journals" (Elaine Surbeck, Eunhye Park Han, and Joan E. Moyer); (15) "Using the Computer as a Reading Portfolio" (Judy Hetterscheidt, Lynn Pott, Kenneth Russell, and Jakke Tchang); (16) "Teachers' Stories: Our Ways of Knowing" and (Mary Renck Jalongo). "Section 5: Significant Outcomes" includes the following essays: (17) "Learning from an Outcomes-Driven School District" (Tom Rusk Vickery); and (18) "Thinking: How Do We Know Our Students Are Getting Better at It?" (Arthur L. Costa). Each essay contains references. (Contains 10 figures.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

K–College

ED 425 194

A Collection

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

L. Stiegman

2A TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM029264

EDITED BY KAY BURKE

2

Authentic Assessment

A Collection

Edited by Kay Burke

SkyLight
TRAINING AND PUBLISHING, INC.
Arlington Heights, Illinois

Authentic Assessment: A Collection

Published by IRI/SkyLight Training and Publishing, Inc.
2626 S. Clearbrook Dr., Arlington Heights, IL 60005
800-348-4474 or 847-290-6600
FAX 847-290-6609
iriskylight.com
http://www.iriskylight.com

Editors: Julia E. Noblitt, Linda A. Perez, Erica Pochis
Book Designer: Michael A. Melasi
Production Coordinator: David Stockman

© 1992 IRI/Skylight Training and Publishing, Inc.
All rights reserved.
Printed in the United States of America.

LCCCN 92-62484
ISBN 0-932935-51-6
0564D-5-97CHG
Item Number: 0890
06 05 04 03 02 01 00 99 98 97 15 14 13 12 11 10 9 8 7 6 5 4

Contents

Introduction 5

SECTION 1: ASSESSING ASSESSMENT 9

- The Nation's Report Card Goes Home: Good News and Bad About Trends in Achievement 13
  by Robert L. Linn and Stephen B. Dunbar
- Budgets, Politics, and Testing 31
  by Chris Pipho
- Why We Need Better Assessments 37
  by Lorrie A. Shepard

SECTION 2: TESTING AND THOUGHTFULNESS 49

- Testing and Thoughtfulness 53
  by Rexford Brown
- Project Spectrum: An Innovative Assessment Alternative 59
  by Mara Krechevsky
- Teaching to the (Authentic) Test 69
  by Grant Wiggins

SECTION 3: ALTERNATIVES TO TESTING 83

- Searching for Alternatives to Standardized Tests: Whys, Whats, and Whithers 87
  by Walter Haney and George Madaus
- Assessment Literacy 101
  by Richard J. Stiggins
- Performance Assessment: The Message from England 115
  by Desmond L. Nuttall
- But...the Curriculum 123
  by Carolyn Mamchur

- Linking Restructuring to Authentic Student Achievement 133
  by Fred M. Newmann
- What Teachers Learn from "Kid Watching" 149
  by Richard VanDeWeghe
- Assessment: A Thoughtful Process 157
  by Steven Ferrara and Jay McTighe

SECTION 4: JOURNALS AND PORTFOLIOS 179
- Assessing Reflective Responses in Journals 181
  by Elaine Surbeck, Eunhye Park Han, and Joan E. Moyer
- Using the Computer as a Reading Portfolio 187
  by Judy Hetterscheidt, Lynn Pott, Kenneth Russell, and Jakke Tchang
- Teachers' Stories: Our Ways of Knowing 191
  by Mary Renck Jalongo

SECTION 5: SIGNIFICANT OUTCOMES 201
- Learning From an Outcomes-Driven School District 205
  by Tom Rusk Vickery
- Thinking: How Do We Know Students Are Getting Better At It? 213
  by Arthur L. Costa

Authors 221
Acknowledgments 225
Index 229

Authentic Assessment
A Collection

We are interested in observing how students produce knowledge rather than how they merely reproduce knowledge. A critical characteristic of intellectual ability is not only having information, but knowing how to act upon it.—Arthur L. Costa

What is authentic assessment? Is it just another "buzz word" of the 1990s that falls under the ubiquitous umbrella of restructuring? If it is a new direction in education, does that mean that all the prior assessment used in schools for hundreds of years has *not* been authentic? Have we been using bogus ways to assess learning? Is it possible that teachers have been subjecting their students to unauthentic means of assessments; therefore, all of the grades have been inaccurate?

Not to worry! Skillful teachers have been utilizing *good* assessment techniques successfully long before the term authentic assessment appeared in educational journals or on the national seminar circuit.

Skillful teachers do not rely solely on standardized tests, commercial tests, ditto worksheets, or the questions at the end of the chapter to show them what their students can do. They construct meaningful tasks that require students to synthesize information, apply what they've learned, and perform or demonstrate their understanding of the material according to specific criteria. In short, they motivate students to learn, and they require students to perform tasks or create products to demonstrate their understanding of the process. They teach students how to "produce" knowledge not "reproduce" knowledge. Unfortunately, the tremendous emphasis on standardized testing has clouded our vision of what's really essential for

life, and we rely more and more on multiple-choice tests and paper-and-pencil tests to measure growth.

Authentic assessments are real-life tools that reflect skills necessary for learning and for life. Like the Olympic rings, assessment tools must be linked to be effective. Students need to construct learning for themselves, link it to prior knowledge, and bridge it to other subjects and real experiences. They need to be able to see the connections between what they are studying in school and what they are experiencing in life.

Educators must practice Costa's principle of "selective abandonment" when it comes to curriculum realignment. Factual knowledge is doubling approximately every eighteen months; therefore, teachers cannot practice the "pour and store" philosophy of learning where the student is the empty vessel into which the teacher pours content with very little retention taking place. Moreover, much of the "knowledge" base will soon become obsolete. Teachers must teach students how to learn for themselves. As Costa says, "Intelligence is knowing what to do when you don't know what to do," and students must learn how to make learning relevant for themselves. Tests cannot always measure or predict students' abilities to apply what they learn in school to life.

What do we want our students to know twenty-five years from now? Memorizing discrete bits of knowledge and practicing mundane drill and skill activities will not be as important as being able to solve problems, communicate with others, make decisions, think critically and creatively, and to participate in the global community as lifelong learners.

If our significant outcomes include helping students survive in the 21st century, then they must also include opportunities to practice these skills and processes and perform meaningful tasks in the classroom. If we want students to transfer classroom learning to life, we must integrate our learning and help them make connections with other subjects and life experiences.

Authentic assessment allows students to practice and to perform meaningful tasks that mirror life outside the school room walls: tasks such as writing, giving speeches,

creating projects, working with others, engaging in problem-solving activities, making decisions and choices about what and how they learn, and using clear criteria to evaluate their own work. Teachers can help by providing opportunities for meaningful interactions with other students, introducing them to multiple resources and speakers, allowing them to reflect on their own work and all formative and authentic assessments.

Teachers can use a wide variety of formative assessments to foster their students' success despite individual learning styles, learning disabilities, and language or behavioral problems. Teachers can also monitor assignments to determine who needs help and who needs enrichment. More emphasis can be placed on genuine and formative tasks rather than on summative evaluation tasks such as unit tests that signal the end of a learning segment and the end of a student's chance to learn it. If it is important enough to learn, it is important enough for *all* students to learn.

We need to expand our repertoire of teaching strategies and formative authentic assessment tools to help students interact with the material, the teacher, and other students in order to enhance higher-order thinking skills and cooperation. The following articles describe how teachers can best accomplish these goals. Authentic assessment cannot stand alone. In order for the assessment to be authentic, the content has to be authentic. Also, the outcomes that drive the curriculum must be authentic. Significant student outcomes, meaningful content, and quality instruction must be components of the instructional package in order for the assessment to be valid. Moreover, educators need to develop "assessment literacy" so they can recognize and deliver meaningful content. We cannot afford to give our students anything but quality instruction and that quality can only be enhanced by the use of authentic tools to assess students' cognitive, social, and emotional development.

Assessing Assessment

State assessment programs are entering a period of uncertainty. They are either on the verge of mammoth transformations or about to freeze up in a policy-making gridlock.–Chris Pipho

Standardized tests and teacher-made tests are too often the yardsticks by which our schools are measured today. The number of tests given in our nation's schools has increased steadily over the last twenty years. But do these tests really reflect what is happening in the classroom? Property taxes, educational legislation, and even elections are influenced by the publication of the "Nation's Report Card," the results of standardized test scores by schools, districts, and states. Important decisions involving placement of students, curriculum changes, budget revisions, and personnel placement are based on the interpretation of such test results; these results, however, are often a subject of contradiction and controversy. Some teachers are being evaluated by the results of their students' standardized tests and *America 2000* calls for providing merit pay for teachers who raise their students' test scores.

Many people measure educational progress and success in schools by the yearly publication of these standardized test results. We are a test-driven educational system. Unfortunately, many people feel that the only way to monitor student growth and to evaluate the effectiveness of restructuring is to continue to rely on standardized tests as the yardstick to measure reform.

Robert L. Linn and Stephen B. Dunbar state that test results have long been used to judge educational achievement in the United States. Moreover, when the public is

actively seeking educational reform, it turns to test scores to measure the results.

Several states and testing organizations, however, are revising their testing programs to reflect an emphasis on problem-solving, higher-order thinking, and performance skills. But the fact remains that too many questions on standardized tests consist of lower-level recall questions that test prior knowledge. Lorrie A. Shepard states that "In the United States today, standardized testing is running amok." Furthermore, "test fever" has resulted in some of the following practices and assumptions that Shepard and others have found:

- Teachers spend an inordinate amount of instructional time preparing students for standardized tests. They "teach for the test" by concentrating on discrete facts and rote "drill and skill" exercises.
- Teachers often neglect social studies, science, and other subject areas to concentrate on math and language arts skills because they are the major areas covered on standardized tests.
- Students are often grouped or tracked as early as first grade based on their standardized test scores.
- Low test scores damage the self-esteem of young people because they are often relegated to the "Blackbird" reading group or to the basic skills or remedial tracks in school.
- Teachers in some areas receive merit pay for raising students' test scores.
- Superintendents and principals are sometimes hired and fired based on their district or school test scores.
- Real estate agents boast about the higher test scores in school districts. Higher scores can mean higher property values.
- The "Nation's Report Card" publishes test scores and rarely is the socioeconomic status or the per pupil money allotted to education in the district mentioned.
- Commercial tests cost a tremendous amount of money.
- Declines in test scores on the Scholastic Aptitude Test (SAT) and American College Test (ACT) are often reported without reference to the large expansion of the test pool that started in the late 1960s and continues today.

• Cheating scandals in districts where pressure is placed on teachers and administrators to score above their counterparts and/or above the national average have increased.

Standardized tests scores are just one component in the educational process. Shepard concurs when she states that "In today's political climate, tests are inadequate and misleading as measures of achievement. Assessment tasks should be redesigned—indeed, are being redesigned—to more closely resemble real learning tasks."

The Nation's Report Card Goes Home:

Good News and Bad About Trends in Achievement

by Robert L. Linn and Stephen B. Dunbar

When data from different sources conflict, the truth is often to be found somewhere in the middle, Messrs. Linn and Dunbar point out. Some such truths are evident from the surveys of achievement trends discussed here.

TEST RESULTS have long been used to judge educational achievement in the U.S. Public interest in test scores appears to be highest during periods when the public is actively seeking education reforms.[1] Not only are test results used as indicators that reform is needed, but they have also come to be viewed as a major mechanism of reform. Thus it should not be surprising that current interest in test scores is at an all-time high.

There appears to be an abundance of data from which to judge educational achievement. However, the interpretation of the results is quite variable.

Each fall headlines report whether average scores on the American College Testing (ACT) Program's college entrance test and the College Board's Scholastic Aptitude Test (SAT) have gone up or down for the nation, the state, and the local district. The press also calls attention to differences among the scores of blacks, Hispanics, and whites and to any change in the size of these gaps. The periodic release of results from the National Assessment of Educational Progress (NAEP), now known as the "Nation's Report Card," attracts similar attention. These national reports are supplemented by reports of the results

From *Phi Delta Kappan*, Vol. 72, no. 2, p. 127-133, October 1990. Reprinted with permission.

of state- and district-level testing programs and by occasional reports of international assessments.

All in all, there appears to be an abundance of data from which to judge educational achievement. However, the interpretation of the results is quite variable, sometimes apparently contradictory, and frequently subject to controversy. Within the past year or two citizens of most states were apt to read reports indicating that students in their state or district continued to score above the national average and had scored better than their counterparts in the previous year. Those same citizens were also likely to come across reports indicating that scores on the ACT and on the SAT exhibited almost no change, an outcome that U.S. Secretary of Education Lauro Cavazos called "disheartening."

From the NAEP, the public would have heard both good news and bad. The details of the NAEP reports vary somewhat by the content area of the assessment and by the age of the students, but, by way of overall summary, substantial gains in achievement were recorded by black and Hispanic students during the 1980s, while the achievement of white students changed relatively little. The absolute level of achievement, especially when put in the context of an international comparison released in January 1989, leaves much to be desired.

Because of the changing pool of test-takers, the SAT results clearly give a distorted picture of trends in student performance for the nation as a whole.

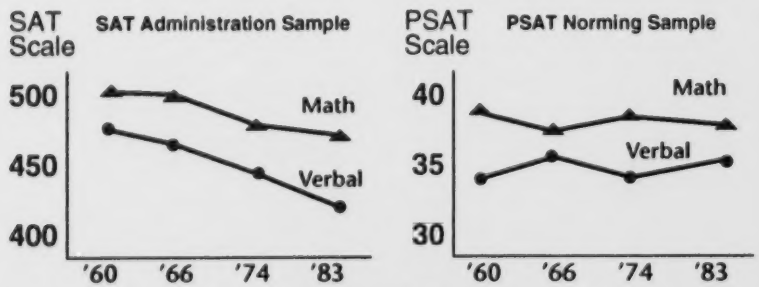
Some of the apparent discrepancies in messages being sent by test results can be reconciled by taking into account differences in who takes the tests (e.g., self-selected college-bound students or a cross section of high school or elementary school students), in the content of the tests, and in the conditions under which the tests are administered. For example, a substantial part of the highly publicized decline in SAT scores that occurred during the late 1960s and throughout the 1970s can be explained by an expansion of the pool of high school students who were applying to college and taking the test during this period. This is evident from the display in Figure 1 of means for the SAT and for the Preliminary Scholastic Aptitude Test (PSAT).[2]

The PSAT results are based on the national norming studies that were conducted in 1960, 1966, 1974, and 1983 and thus are results for comparably defined groups of students. The SAT results, on the other hand, are based on the groups of students who happened to take the test in those four years and hence are not comparable. Because of the changing pool of test-takers, the SAT results clearly give a distorted picture of trends in student performance for the nation as a whole.

Changes in performance in one content area obviously need not match those for another. A gain in math or science achievement, for example, might occur during the same period in which there was a decline in achievement in civics or history. Indeed, it is often suggested that high-stakes testing in limited content areas may lead to an overemphasis on those areas at the expense of content areas not covered by the high-stakes tests.[3]

More subtle—and more difficult to explain to the general public—are changes within a content area that affect the meaning and importance of performance on specific test items. As any teacher knows, changes in what is taught greatly affect the specifics of what students know and can do. For example, at the turn of the century it was not unusual for tests required for graduation from the eighth grade to emphasize arithmetic calculation (including such problems as the hand calculation of square roots) and the memorization of names, dates, and places in history. Questions that assess such knowledge and skills would have little value for judging the comparative success

Figure 1: Average SAT/PSAT Performance by Year



of students who have studied a history curriculum that is issues-oriented and a mathematics curriculum that emphasizes conceptual understanding and problem solving and that makes heavy use of calculators.

Differences in the purposes of testing and the conditions of testing can also influence results. The NAEP does not report scores for individual students, schools, or even school districts. Indeed, obtaining state-level results from the NAEP became possible only recently.[4]

These circumstances are quite different from the circumstances surrounding typical administrations of achievement tests by states or districts. In such testing programs, not only are individual student scores reported, but, in the case of high-stakes testing, schools and individual teachers are also held accountable for the results. Students' motivation to do well when the scores are sent home and teachers' motivation to teach to the content of the test when they are being evaluated in terms of their students' scores are quite different from the conditions that apply during an administration of the NAEP.

The average achievement of blacks and of Hispanic students is substantially higher now than a decade ago.

Differences in test security can also partly explain why state and district reports of test results have generally painted a much more positive picture of progress in educational achievement than is provided by the NAEP. There is good reason to believe that the more positive picture derives, in part, from such factors as the exclusion of some students from testing and students' familiarity with the specific content of the tests (since the same forms of the tests are used year after year).[5] There is also reason to believe that the testing conditions in the administration of the low-stakes NAEP may result in an overpessimistic view of student achievement.

After taking into account these differences in the populations of test-takers, in the content of the tests, and in the conditions of test administration, there still remains a mixture of good news and bad. As we will show below, there has been some progress in achievement, most clearly at the low end of the scale. Progress at the high end of the

scale is more difficult to ascertain because the results are equivocal. Clearly, the good news is that the average achievement of blacks and of Hispanic students is substantially higher now than it was a decade ago. But the gap between the achievement of minority students and that of white students remains large.

NAEP RESULTS

Perhaps the best single source of information about trends in educational achievement is the National Assessment of Educational Progress. The NAEP avoids the problem of changing pools of test-takers, which plagues the SAT and the ACT, by assessing nationally representative samples of in-school 9-, 13-, and 17-year-olds. While the NAEP content is not designed to match any particular curriculum, it is based on a consensus that attempts to insure that it reflects "the thinking of a wide variety of individuals—state and district curriculum specialists, teachers, school administrators, researchers, parents, concerned citizens, officials, and business people."[6] The NAEP is administered under uniform conditions that protect the security of the tests. Finally, the NAEP includes assessments in several key content areas, including reading, writing, mathematics, science, civics, history, computer literacy, and geography.

Reading. Because of the central role of reading in all school subjects and in the development of an educated citizenry, reading is the most frequently assessed content area in the NAEP. Results are now available from the five assessments conducted between 1971 and 1988. Although the average 9-year-old and the average 17-year-old in 1988 had significantly higher reading performance than their counterparts in 1971, the overall increases were relatively small, and the performance of 13-year-olds has remained essentially unchanged since 1971. Furthermore, the gains that were achieved by 9-year-olds occurred between 1971 and 1980, and the average achievement levels in 1984 and 1988 were slightly lower than the level obtained in 1980.[7]

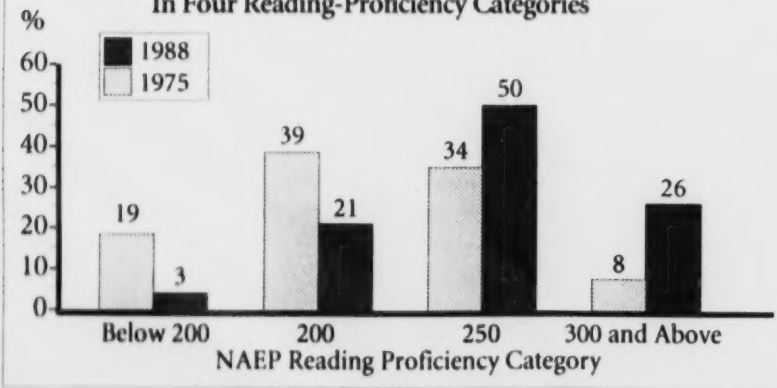
To go beyond this relatively sterile global summary of the results, it is necessary to see how the best and poorest readers in 1988 compare to their counterparts in

earlier years. It is also essential to get a sense of what the scores mean in terms of what students can and cannot do. Finally, it is important to consider separately the performance of minority students, who have traditionally been poorly served by schooling and consequently have recorded lower average levels of achievement. We will go back only to 1975 for this discussion, because the 1971 assessment used a different way of breaking down the sample into racial and ethnic groups than was used in subsequent assessments.

The NAEP reading scale is anchored with example questions and descriptions at five levels. We will focus on the middle three of these: scores of 200, 250, and 300. The NAEP refers to these levels as basic, intermediate, and adept, but such labels do not convey an overall sense of the reading proficiency of students scoring at these levels.

According to the standard NAEP descriptions,[8] students who score at the 200 level "have learned basic comprehension skills and strategies and so can locate and identify facts from simple informational paragraphs, stories, and news articles." They can also "combine ideas and make inferences from short, uncomplicated passages." Students who score at the 250 level "can search for, locate, and organize the information they find in relatively lengthy passages and can recognize paraphrases of what they have read. They can also make inferences and generalizations about main ideas and about an author's pur-

Figure 2: Distribution of Black 17-Year-Olds In Four Reading-Proficiency Categories



pose." Students who score at the 300 level "can understand complicated literary and informational passages, including material about topics they study at school. They can also analyze and integrate less familiar material and provide reactions to and explanations of the text as a whole."

The anchor points of the NAEP scale are chosen in such a way that a student who scores at the 250 level is highly likely to answer correctly a question at the 250 level or below and is unlikely to answer correctly a question at the 300 level. So when it is said that 50% of the nation's 13-year-olds performed at the 250 level, it doesn't follow that none of the remaining 50% could answer any of the 250-level questions—only that they were unlikely to answer correctly any given question at that level.

Insuring that all students progress and that the achievement gap between minority and majority students continues to close must be a national priority in the years to come.

It is startling to realize that 19% of 17-year-old black students had levels of reading achievement below 200 in 1975 and that only 8% of 17-year-old black students scored at the 300 level or above. The good news (presented in Figure 2) is that by 1988 the percentage of black 17-year-olds scoring below 200 had decreased from 19% to 3%, while the percentage scoring at 300 or above had increased from 8% to 26%.[9] These results, together with comparable improvements for Hispanic 17-year-olds between 1975 and 1988, show that the whole distribution of reading proficiency scores of black and Hispanic 17-year-olds shifted to higher levels. That is, not only were average scores higher, but the levels achieved by the top and bottom readers within both of these minority groups increased in absolute terms and in comparison to the performance of their white peers.

White students, on the other hand, made only small gains during this period. But white students constituted the great majority of students in the two highest categories (86% at or above 250 in 1975, rising to 89% in 1988). Since the comparable figures for black and Hispanic 17-year-olds in 1988 were 76% and 73% respectively, it is also clear that there is still a substantial gap between the

reading proficiency of the average black or Hispanic 17-year-old and that of the average white 17-year-old. Insuring that all students progress and that the achievement gap between minority and majority students continues to close must be a national priority in the years to come.

Figure 3: NAEP Trends in Reading for 17-Year-Olds At Three Performance Levels

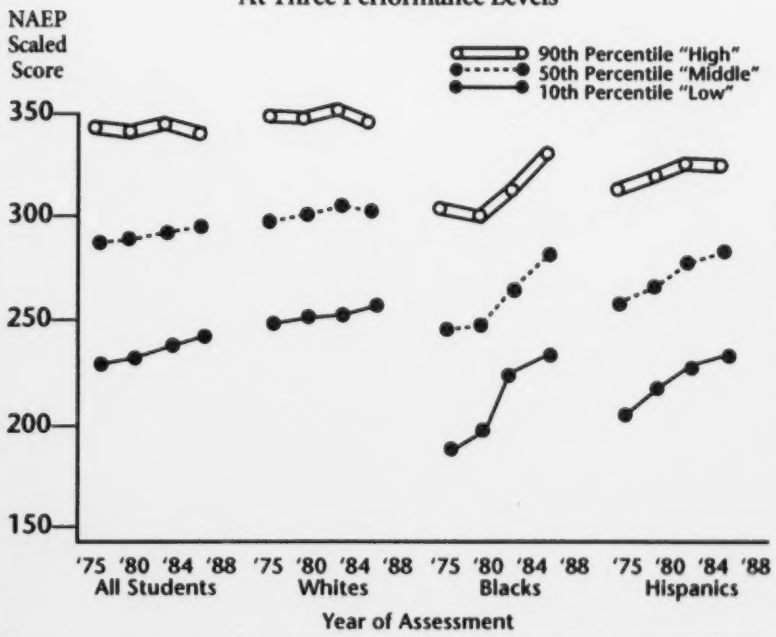


Figure 3 provides a comprehensive picture of the progress of 17-year-olds over the last four assessments in terms that are more familiar outside the context of the NAEP and that can be more readily compared to other data sources.[10] The scores achieved by high-, middle-, and low-scoring students within each of the three major racial/ethnic groups and for the total group are plotted for each of the last four assessments. *High* is here defined as the 90th percentile for a given group and year: that is, the score below which 90% of the students in a given group are found. *Low* is defined as the 10th percentile for the group, and *middle* is defined as the group median.

There is a marked contrast between the trend lines for minorities and those for whites at all three levels. Black students—and to their lesser extent Hispanic students—have demonstrated unequivocal gains in reading comprehension, especially since the 1980 assessment. White students, on the other hand, have an achievement pattern much like that of the total sample: a slight downward trend among high-achieving students, coupled with only modest increases among middle- and low-achieving students. These and similar results at the other two age levels and in other content areas have led to statements in the NAEP reports that the gains have generally been at the more basic levels rather than on tasks requiring higher-order thinking. As we discuss below, the continued narrowing of the gap between minority and nonminority students—rather than the overall trends in the total sample—is bolstered by evidence from other sources of achievement data for the nation.

The trends in reading achievement for 9- and 13-year-olds are more mixed than those for 17-year-olds. Between 1975 and 1980, black, Hispanic, and white 9-year-olds at all achievement levels recorded gains. The performance achieved by the highest-scoring students within each of these groups was also higher in 1988 than in 1980. However, the performance achieved by average- and low-scoring 9-year-olds in 1988 remained relatively constant or in some cases (e.g., low-scoring white students and average-scoring black students and white students) declined slightly.

U.S. seniors ranked 14th in advanced algebra, 12th in elementary function/calculus, and 12th in geometry.

At age 13, the performance of white students remained essentially unchanged across the four assessments, while the performance of black and Hispanic students improved. The improvement for Hispanic students occurred at all achievement levels—but at less dramatic rates than those posted by 17-year-olds.

Mathematics and science. Before considering the specifics of trends in achievement in mathematics and science, it is worthwhile to place the issue in the context of

national needs and international comparisons. The mathematical needs of the U.S.—indeed, of any advanced technological society—are eloquently and forcefully laid out in *Everybody Counts*, a recent report from the National Research Council. As that report makes clear, mathematics "is the key to opportunity and careers."[11] The importance of mathematics, science, and technology and the need to restructure U. S. precollege education in these areas are also discussed in *Science for All Americans*, a report from Project 2061 of the American Association for the Advancement of Science.[12]

The nation can no longer afford to turn children away from mathematics and science, and the nation can no longer live with the myth that these subjects require "special ability, which most students do not have."[13] One legacy of the myth is that U.S. citizens make up a shrinking fraction of the nation's graduate school enrollment in science and mathematics. Between 1970 and 1987, for example, there was a 50% decline in the number of doctorates in the mathematical sciences awarded to U.S. citizens; by 1987 the majority of new doctoral decrees in mathematical sciences awarded by U.S. universities went to citizens of foreign countries.[14]

In 1981-82 the Second International Mathematics Study (SIMS) obtained achievement test results and a good deal of information about content coverage for eighth-graders from 20 countries (in the case of Japan and Hong Kong, seventh-graders). Test results were also obtained for students from 15 countries who were in their last year of high school. Of the 20 countries participating at the eighth-grade level, U.S. students ranked 10th in arithmetic, 12th in algebra, 16th in geometry, and 18th in measurement. Of the 15 countries participating in the assessment in the senior year of high school, the U.S. ranked 14th in advanced algebra, 12th in elementary function/calculus, and 12th in geometry. Even after taking into account differences in the percentage of students retained through the last year of high school, the level of achievement of U.S. 12th-graders compared unfavorably to that of students in such countries as Japan, Hong Kong, Sweden, Israel, Finland, New Zealand, and England and

was similar to that achieved by students in Thailand and Hungary.[15]

Another international assessment involving fewer countries was conducted in mathematics and science in 1988. Samples of 13-year-olds from Ireland, Korea, Spain, the United Kingdom, four Canadian provinces (British Columbia, New Brunswick, Ontario, and Quebec), and the U.S. were assessed using math and science exercises from the 1988 NAEP. Since separate assessments were conducted in French and in English for three of the Canadian provinces, seven separate sets of test results were reported for Canada. Thus 12 populations of students were compared (five countries plus seven province/language groups). U.S. students ranked either last or next to last among these 12 groups on each of five types of problems in mathematics and next to last in each of three categories of science problems.[16]

International rankings are subject to many problems, including sampling, translation from one language to another, differences in retention policies and differences in curriculum timing. All can cause difficulty in interpreting comparisons such as those just presented. Despite these limitations, the results for the U.S. aroused concern. The 1988 results are particularly sobering in that the exercises were all developed specifically for U.S. students as part of the NAEP. Thus the poor showing of U.S. students cannot be dismissed because the exercises were more in line with the curriculum of another country. In any event, these international results provide a larger context for considering trends in achievement in math and science.

Since 1978 black students have outdistanced their peers in other racial/ethnic groups in the degree of improvement on the NAEP mathematics tests.

As it does with the reading scale, the NAEP identifies levels of performance in mathematics, anchored at five points on the scale by verbal descriptors and exemplary questions. The three middle levels have been designated beginning skills and understanding (level 200), basic operations and beginning problem solving (level 250), and moderately complex procedures and reasoning (level 300).

> Students who perform at or above level 200 are developing a greater range and depth of basic mathematical skills; however, their use of these skills is still imperfect and relatively inflexible. It can be inferred that learners at this level would have difficulty with reasoning that requires more than simple numerical computation.... Students who perform at or above level 250 on the proficiency scale have developed a surface understanding of the four basic operations and are beginning to acquire more complex reasoning skills.... Students who perform at or above level 300 demonstrate sophisticated numerical reasoning and are beginning to draw from a wider range of mathematical skill areas, including algebra and geometry.[17]

As in reading, some students at level 200 will answer correctly some questions that exemplify level 250 skills, though they would be unlikely to do so. Conversely, some students at level 300 may answer incorrectly questions that exemplify level 250 skills, though they too would be unlikely to do so.

The only stable trends in the percentage of students at each age level who attain a given proficiency occurred at the 200 and 250 levels. The direction of change at these levels has been positive since the 1978 assessment. While achievement described in this manner has improved, the NAEP notes in *The Mathematics Report Card* (which was based on the 1986 NAEP results) that only about 21% of the 9-year-olds achieved the level of basic operations and beginning problem solving (level 250), that only about 16% of the 13-year-olds reached the level of moderately complex procedures and reasoning (level 300), and that 6.4% of the 17-year-olds were proficient at skills that were described as "multi-step problem solving and algebra" (level 350). This last finding, coupled with the lack of any upward trend in high-level achievement, prompted much concern in the report of the 1986 assessment over the lack of progress in complex mathematical reasoning.

NAEP data can be used to trace overall trends in mathematics achievement from 1973 through 1986. In general, these data show only minor changes over the 13-year period. For example, the performance of 9- and 13-

year-olds rose by a modest (albeit statistically significant) amount. In contrast, the scores of 17-year-olds have shown a net decline since 1973, although improvements since 1982 appear to be stable. As was true for reading, however, the most dramatic changes in the NAEP mathematics findings are the consistent gains made by blacks during the past three assessments. Since 1978 black students have outdistanced their peers in other racial/ethnic groups in the degree of improvement on the NAEP mathematics tests.

The average achievement in the NAEP science assessment increased at all three age levels between 1982 and 1986.[18] But that increase came after a relatively steady decline among 17-year-olds across the four assessments that were conducted between 1969 and 1982, and the subsequent increase for this age group has not been large enough to make up for the earlier decline. The performance of 17-year-olds in 1986 returned to where it had been in 1977 but remained below the level achieved by 17-year-olds in 1969 and in 1973. For 9- and 13-year-olds, the trend line for average achievement, while showing some slight ups and downs between 1970 and 1986, has been nearly flat.

What progress has been made in the sciences and mathematics must be carefully nurtured to see that the gains of the Eighties continue and even accelerate in the Nineties.

At all three age levels there was a narrowing of the gap between the average science achievement of white students and that of black students—though the relative gains were less dramatic than those posted in reading by black 17-year-olds. The gap in average achievement between Hispanic and white students also narrowed slightly between 1977 and 1986 at ages 9 and 13, but not at age 17. Despite such gains, there remains much to be done at national, state, and local levels to eliminate the still-substantial group differences in achievement for minority students in both science and mathematics.

ADDITIONAL DATA

Besides the NAEP and the college admissions testing programs, data on trends in achievement are available

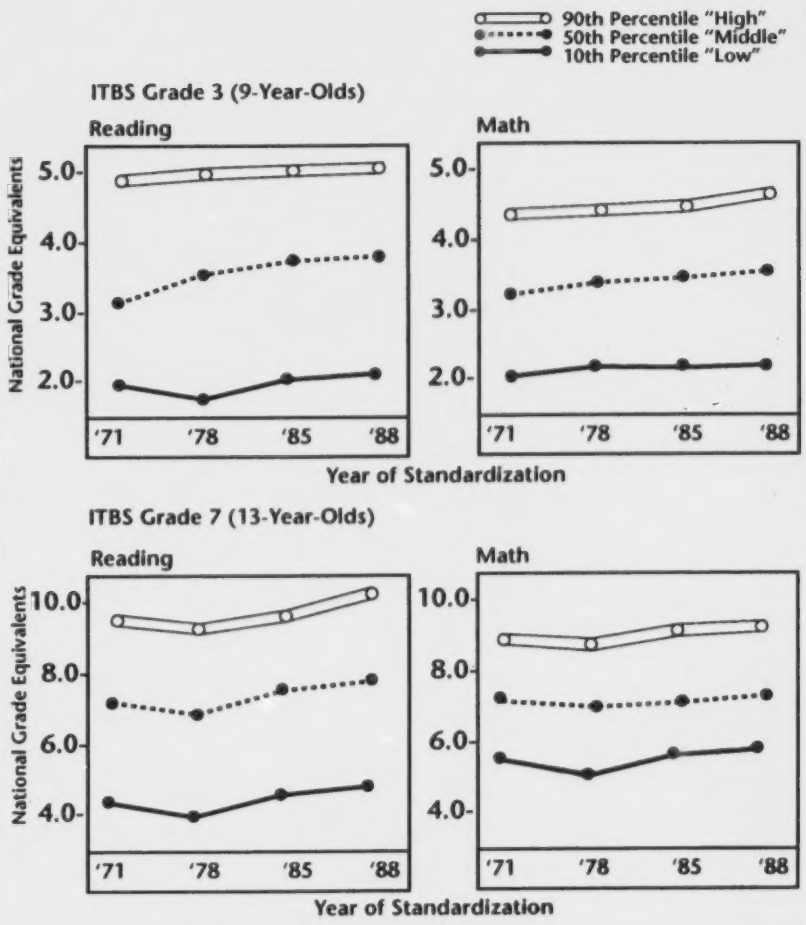
from nationally standardized achievement tests and from the records kept by various statewide testing programs. Perhaps the most complete historical data on educational achievement, analyzed in a recent report from the Congressional Budget Office,[19] come from the Iowa Testing Programs in the form of trend lines for both the nation and the state of Iowa. A renorming of the Iowa Tests of Basic Skills (ITBS) during the 1987-88 school year permits us to examine trends in a nationally representative sample tested under standardized conditions during roughly the period of time covered by the NAEP reading and mathematics assessments.

The trend lines in Figure 4 show a somewhat different picture of achievement in reading and mathematics than was apparent with the NAEP. The performance of a nationally representative sample of third-graders used to standardize the ITBS has shown steady improvement since 1971 and, in contrast with the NAEP, has shown gains at even the highest achievement levels. Although the well-publicized decline in achievement test scores continued well into the Seventies for the seventh-graders, the ITBS data indicate the decline to have ended for seventh-graders by 1978. In reading and mathematics, the rise begun in 1978 continued into the most recent renorming of the ITBS. Indeed, national achievement measured by the ITBS now exceeds the highest point registered before the score decline of the 1960s. As in the NAEP, the more dramatic gains are observed among the lower-achieving students; however, increases at the upper ends of the score distributions are noteworthy in their consistency across all content areas.

The performance of minorities has also shown marked improvement since the 1977-78 standardization of the ITBS. By the time of the 1984-85 standardization, for example, black students in grade 5 had made up about 30% of the gap between their achievement and that of their white peers in the verbal skills measured by the ITBS (vocabulary, reading, language usage, spelling, and so on). However, no consistent narrowing of the achievement gap was registered in quantitative skills (math concepts, problem solving, and computation).

A recent University of Iowa study sheds additional light on the NAEP findings that progress in higher-order skills has been limited. The Iowa study examined the performance of comparable samples of 9-, 13-, and 17-year-olds on test questions in a given subject that were classified as measuring either higher- or lower-order skills.[20] Between 1979 and 1985, performance increases were found on both types of questions. However, contrary to the interpretations of the NAEP, the increases were

Figure 4: ITBS Trends at Three Performance Levels



slightly larger for the complex, higher-order skills. The authors of this study argued that there was little empirical evidence that increases in achievement during the period were limited to relatively mechanical tasks.

* * *

When data from different sources conflict, the truth is often to be found somewhere in the middle. Some such truths are evident from the surveys of achievement trends discussed above. The gains among low achievers and minorities, in particular, are clear and consistent for all sources of data. What trend data say about the performance of students in, say, the top 10% of their classes is less clear, as are the correct inferences for policy making that can be drawn from any of these apparent trends. Other writers have urged that the education community not presume that a given reform effort will have a desired effect on achievement when the causes of achievement trends are multifarious.[21]

We urge similar caution. But we also suggest that the most recent round of report cards sends several messages to the public.

Reports of achievement trends can be very much influenced by the particular definition of achievement that a test represents and by the conditions under which the test is administered. These factors no doubt account for some of the discrepancies detailed in this article. However, it is through a diversity of data on achievement trends that the most accurate, albeit more complex, portrait of educational achievement can be painted. To maintain this diversity is to broaden our understanding of the complexities of educational development and to increase the likelihood of detecting any unanticipated effects of reforms.

Despite the potential flaws of international surveys of achievement, the perspective they provide in mathematics and science suggests that efforts be made in these areas to stimulate and maintain the interests of young people at all levels of achievement and in all subgroups of the student population. What progress has been made in the sciences and mathematics must be carefully nurtured to see that the gains of the Eighties continue and even accelerate in

the Nineties. We also have reason to nurture what have come to be labeled higher-order reasoning and problem-solving skills, in both the verbal and the quantitative domains. It is important that such skills be properly represented in the definition of achievement that is embodied in all forms of assessment, so that progress in this area, too, can be accurately measured.

The authors wish to thank H. D. Hoover for valuable insights and data pertaining to this article.

NOTES

1. Daniel P. Resnick. "History of Educational Testing," in Alexandra K. Wigdor and Wendell R. Garner, eds., *Ability Testing: Uses, Consequences, and Controversies, Part 2*. (Washington, D.C.: National Academy Press, 1982), pp. 173-94.
2. *ETS Policy Notes: News from the ETS Policy Information Center*, vol. 2, October, 1989, p. 6; and Thomas F. Donlon, ed., *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests* (New York: College Entrance Examination Board, 1984), p. 176.
3. Edward Haertel, "Student Achievement Tests as Tools of Educational Policy: Practices and Consequences," in Bernard R. Gifford, ed., *Test Policy and Test Performance: Education, Language, and Culture* (Boston: Kluwer Academic Publishers, 1989), pp. 25-50.
4. P.L. 100-297, Part C, provides for the collection and reporting of state-level NAEP results on a trial basis in 1990 and 1992, with voluntary participation.
5. Robert L. Linn, M. Elizabeth Graue, and Nancy M. Sanders, "Comparing State and District Test Results to National Norms: The Validity of Claims that 'Everyone Is Above Average,'" *Educational Measurement: Issues and Practice*, in press.
6. *Reading objectives: 1986 and 1988 Assessments, National Assessment of Educational Progress* (Princeton, N.J.: Educational Testing Service, 1987), p. 5.
7. Ina V. S. Mullis and Lynn B. Jenkins, *The Reading Report Card, 1971-88: Trends from the Nation's Report Card* (Princeton, N.J.: National Assessment of Educational Progress and Educational Testing Service, 1990).
8. Ibid., pp. 26-30.
9. Ibid., pp. 63-64.
10. Ibid., pp. 62-65.
11. *Everybody Counts: A Report to the Nation on the Future of Mathematics Education.* (Washington, D.C.: National Academy Press, 1989).
12. *Project 2061: Science for All Americans* (Washington, D.C.: American Association for the Advancement of Science, 1989).
13. *Everybody Counts...*, p. 10.
14. Ibid., pp. 24-25.
15. Curtis C. McKnight et al., *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective* (Champaign, Ill.: Stipes Publishing Co., 1987); and M. David Miller and Robert L. Linn," International Retention and Achievement Rates," *Journal of Research in Mathematics Education*, vol. 20, 1989, pp. 28-40.

16. Archie E. Lapointe, Nancy A. Mead, and Gary W. Phillips, *A World of Differences: An International Assessment of Mathematics and Science* (Princeton, N.J.: Educational Testing Service, 1989).
17. John A. Dossey, et al., *The Mathematics Report Card: Are We Measuring Up? Trends and Achievement Based on the 1986 National Assessment* (Princeton, N.J.: Educational Testing Service, 1988), pp. 33, 36, 39.
18. Ina V. S. Mullis and Lynn B. Jenkins, *The Science Report Card: Elements of Risk and Recovery—Trends and Achievement Based on the 1986 National Assessment* (Princeton, N.J.: Educational Testing Service, 1988).
19. Daniel Koretz, *Trends in Educational Achievement.* (Washington, D.C.: Congressional budget office, 1986). See also *Achievement Trends in Iowa: 1955-1989* (Iowa City: Iowa Basic Skills Testing Program, 1989).
20. Elizabeth A. Witt, Myunghee Y. Han, and H. D. Hoover, "Recent Trends in Achievement Test Scores: Which Students Are Improving and on What Levels of Skill Complexity?," paper presented at the annual meeting of the National Council on Measurement in Education, Boston, 1990.
21. Daniel Koretz, *Educational Achievement: Explanations and Implications of Recent Trends.* (Washington, D.C.: Congressional Budget Office, 1987).

Budgets, Politics, and Testing

by Chris Pipho

It appears that state testing programs are in for some kind of change. States such as California and Kentucky, and organizations such as the National Governors' Association, are leading the way in the shakeup of the industry.

STATE assessment programs are entering a period of uncertainty. They are either on the verge of mammoth transformations or about to freeze up in a policy-making gridlock. From the national education goals to changes in state policy, demands are being made on the testing establishment that will be difficult to meet. Interpreting the events and forces at work requires a broad field of vision, for many of the actions now being taken are outside the testing establishment. Indeed, that fact may ultimately lead to a stalemate.

The feeling in some state policy-making circles is that a whole new assessment system is needed.

A poll of commercial testing people would probably turn up a variety of responses like the following:

- We are working on tests of actual performance.
- The critics and naysayers don't really understand what a good job we are doing.
- If you want to make such big changes, it's going to take a lot of money and maybe as long as 10 or 15 years.

But the feeling in some state policy-making circles is that a whole new assessment system is needed. Moreover, many policy makers believe that we can get that system

From *Phi Delta Kappan*, vol. 72, no. 2, p. 102-103, October 1990. Reprinted with permission.

only by dumping the present one, reallocating the money, and building a new testing program from the ground up.

Adopting this radical approach will make it impossible to use commercial, off-the-shelf tests or even customized versions of such tests. Depending on how quickly the test writers in state agencies can work, either the situation will lead to a major shakeup of the testing industry—or the whole enterprise will grind to a halt. Business as usual is probably not in the cards, and whoever controls state purse strings will have the advantage.

THE CALIFORNIA BUDGET

Gov. George Deukmejian has line-item veto power over the California budget, and in his last year in office he has used it on education in unexpected ways. After a long budget battle, the governor cut $425 million from Proposition 98 protected funds, another $22 million from nonprotected funds, and $16 million from the state department's budget. One of the largest single programs to be cut is the $12.4 million California Assessment Program (CAP). In his veto message, the governor said that he was setting this money aside "for a new assessment program in California."

State testing officials claim that matrix sampling yields better information while using less student time.

To many observers the governor's action came as a complete surprise. But others interpreted it as a rekindling of old hostilities between Gov. Deukmejian and State Superintendent Bill Honig. Reallocating these and other set-asides would take new legislation, and some speculate that this could not be introduced until 1991. Trying to override the veto was apparently discussed by those opposed to the cuts, but that tactic was ruled out as well. As this column went to press, several bills were being prepared to put the CAP or some new version of it back in place.

Since the governor has not laid out any detailed alternative plan, it remains uncertain how this controversy in California will play itself out. Some sources speculate that business leaders are pushing for individual student test results in order to get a better fix on how well individual students are prepared for college or for work. Currently, the CAP uses a matrix sampling procedure.

State testing officials claim that matrix sampling yields better information while using less student time; they maintain that the scope of the test is 10 times broader than that of traditional tests and that it uses only one-fourth as much instructional time.

The CAP program has always been held in high regard by testing personnel in other states. If the governor does have a new vision of testing in mind, California will be a state to watch.

A NEW PROGRAM IN KENTUCKY

The new reform law enacted in Kentucky last year calls for the state to abolish the use of traditional norm-referenced tests and to adopt a new (and still undeveloped) program of primarily performance-based assessment. At an Education Commission of the States (ECS) steering committee meeting last April, Jack Foster, Kentucky's secretary of education and Gov. Wallace Wilkinson's top education advisor, explained that Kentucky was hoping to use state money earmarked for research and development to produce such a test. Foster asked whether other states would be willing to share the burden of research and development in exchange for benefiting from the results.

Interest in the idea was immediate, and, at the 18th annual ECS conference on assessment held in Boulder in June, ECS asked some of the state testing officials in attendance whether they might be interested in exploring the Kentucky proposal. Officials from more than 10 states, as well as representatives of the U.S. Department of Education, attended an exploratory meeting.

A conference inviting teams of state policy makers and state department of education officials from all 50 states is tentatively scheduled for this fall. The goals of the conference are:

- to develop greater support among state policy makers for authentic performance assessment;
- to help state policy makers understand the issues associated with authentic performance assessment; and
- to develop various consortia of states to accelerate the development and use of authentic performance measures and to resolve the technical and reporting issues.

It is hoped that, by bringing together all the major players from each state, decisions can be made about pooling financial and staff resources and an acceptable division of labor can be agreed upon. Such cooperative efforts will help Kentucky meet its deadline and will keep other states from finding their individual assessment reforms stalled in gridlock.

NATIONAL GOVERNORS' ASSOCIATION

At the summer meeting of the National Governor's Association (NGA), steps were taken to monitor progress toward achieving the national education goals established last year by the NGA and the White House. The governors decided to create a 14-member committee to evaluate each state's progress on implementing the six national goals. The special committee includes six governors (three from each political party), four White House appointees, and four members of Congress (two from the Senate and two from the House of Representatives). The members of Congress will be nonvoting members of the committee.

At the NGA meeting, the governors also hammered out strategies to help states get the job done. Some of these include cutting red tape and reducing regulations that apply to local districts and holding students more accountable for their academic performance. The governors also discussed ways to ease restrictions on federal program expenditures.

Yearly status reports will no doubt have to say something about comparisons of student achievement, and a test of some kind will probably be used to obtain these achievement data.

Plans now call for this independent committee to produce yearly reports evaluating the progress of individual states on the implementation of the six general goals. While a majority of the members of the committee are federal officials or White House appointees, the governors will have a majority of the voting members.

NGA's new chairman, Gov. Booth Gardner of Washington, named Gov. Roy Romer of Colorado to be committee chair. Early speculation favored late September as a formal starting date for the committee's meetings—

about one year after the education summit that started the process of arriving at national goals.

Goal 4 specifies that "U.S. students will be first in the world in science and mathematics achievement," and Goal 3 refers to "demonstrated competency in challenging subject matter." Thus the yearly status reports will no doubt have to say something about comparisons of student achievement, and a test of some kind will probably be used to obtain these achievement data. Since one of the NGA's suggested strategies is to hold students more accountable for their performance, it is not hard to imagine how test scores might come to receive increased emphasis.

It may not be possible to read all of these signs, but some critics still charge that all of this activity smacks of a move toward a national curriculum.

What tests are now available for this sort of assessment? Some assume that the state-by-state comparisons now being developed by the National Assessment of Educational Progress (NAEP) will be pushed onto the fast track. Of course, another option is to come up with an entirely new kind of assessment program at the state level. But getting all the states behind such an effort won't be easy and won't happen overnight. This could place the spotlight squarely on the ECS/Kentucky collaborative effort or perhaps on the commercial testing companies— if the latter have something new to roll out for inspection. If the U.S. is to beat the foreign competition in math and science by the year 2000, some mechanism for making valid international comparisons will also have to be available.

* * *

While all of this has been happening, Roy Truby, the executive director of the National Assessment Governing Board, released a response to an open letter prepared by the National Center for Fair and Open Testing (FairTest). Truby called the FairTest letter part of a "much wider campaign against virtually all forms of testing and educational accountability."

It may not be possible to read all of these signs, but some critics still charge that all of this activity smacks of a move toward a national curriculum. Immediately following the NGA meeting, the president of the National School

Boards Association speculated that local school boards could become the scapegoats, and Samuel Husk, executive director of the Council of the Great City Schools, was quoted in the *New York Times* as saying that the NGA had set up a system by which "everyone can be held accountable." But the NGA committee isn't responsible to anyone.

It would appear that state testing programs are in for some kind of change, and that it will come rather quickly. Will the NGA special committee be the driving force? Will the collaboration among the states produce something useful? Will the NAEP be a part of the mix? Will commercial testing companies come to the rescue? Then again, the whole process could be played out in slow motion, absorbing an endless volume of rhetoric—political and psychometric.

The situation reminds me of a story that ran through the ranks of the U.S. Army in England just prior to D-Day. One G.I. said to another, "Did you hear about Ike's new plan to invade France? He's going to freeze the channel and drive across. Our unit is in charge of getting the ice!"

Why We Need Better Assessments

by Lorrie A. Shepard

Educators should use a variety of assessment measures, make substantive improvements to standardized tests, and remove incentives to teach to the test.

In the United States today, standardized testing is running amok. Newspapers rank schools and districts by their test scores. Real estate agents use test scores to identify the "best" schools as selling points for expensive housing. Superintendents can be fired for low scores, and teachers can receive merit pay for high scores. Superintendents exhort principals and principals admonish teachers to raise test scores—rather than to increase learning. Occasionally school boards issue a mandate that "all students must be above the national norm," which, absurdly, is the same as requiring that 100 percent of students be above the 50th percentile.

Even advocates of high-stakes testing acknowledge that the tests do not cover the full range of important instructional objectives.

LIMITATIONS OF STANDARDIZED TESTS

Current political pressures exaggerate the inherent flaws in paper-and-pencil tests. Why should the public fear that test scores are going up without a real gain in achievement (Cannell 1987)? Because large-scale testing programs are limited by political and practical considerations that undermine their fidelity to important learning outcomes.

From *Educational Leadership*, vol. 46, no. 7, p. 4-9, 1989. Reprinted with permission.

The negotiation of content. First, the content of such tests must be negotiated. Most state-developed tests go through a consensus-building process: teams of curriculum experts and teachers agree to the content of the test. Publishers of standardized tests follow similar procedures; in addition, they do content analyses to ensure that test objectives are well matched to widely used textbooks. These procedures are sensible but they have a homogenizing effect, limiting both the breadth and depth of content coverage. The most imaginative and challenging problem-solving tasks offered uniquely by a single district are likely to be negotiated out of the test content domain. Ironically, the textbooks that commercial test framers look to for verification of coverage are governed by the same sort of consensus model, to ensure market appeal, and are in turn matched to the outlines of standardized tests (Tyson-Bernstein 1988).

The narrowing of content. Test construction is further constrained by the emphasis on basic skills, limiting the "height" as well as the depth and breadth of permissible content. Even advocates of high-stakes testing acknowledge that the tests do not cover the full range of important instructional objectives. Instead, the argument goes, "creative teachers can efficiently promote mastery of content-to-be-tested and then get on with other classroom pursuits" (Popham 1987, p. 682). The extent of this narrowing—that is, whether 10 percent or 50 percent of desirable content is sacrificed—will depend on the particular test and the accountability purpose or market that shaped it. For example, minimum competency or promotional gates tests are even narrower than standardized tests of basic skills.

Multiple-choice formats. Practical considerations also dictate test format in ways that further constrain content. Given the huge numbers of examinees and time limits, test developers use multiple-choice formats to the exclusion of tasks where students would produce or explain a correct answer. Essay tests of subject matter knowledge are rare. Although multiple-choice questions can elicit important conceptual distinctions from students, tests composed entirely of such items do not measure a

respondent's ability to organize relevant information and present a coherent argument. Time limits and the need to survey as much content as possible also preclude problem sets aimed at the interrelation of facets of the same topic.

Other format problems. Similarly, reading passages are strikingly shorter and less complex than the texts students use for daily instruction. Moreover, to avoid confusing students by frequent changes in format, each test uses a small number of item types—despite evidence that students may appear to know a concept or skill when it is measured in one format but not know it if measured another way (Shepard 1988). The effects of these format limitations for the content validity of the test will increase over time, if the same test or a similar test is given each year.

TEACHING TO THE TEST

Despite these numerous drawbacks, the public gives standardized test scores great weight. When the scores have serious consequences—and they often do—teachers will teach to the test. Indeed, it is often the explicit intention of policymakers to force teachers to address essential skills. But teaching to the test cheapens instruction and undermines the authenticity of scores as measures of what children really know, because tests are imperfect proxies even for the knowledge domains nominally covered by the tests; and they also omit important learning goals beyond the boundaries of the test domain.

Teaching to the test cheapens instruction and undermines the authenticity of scores as measures of what children really know.

Researchers have found ample evidence that testing shapes instruction. In one study, for example, elementary teachers reported taking time away from science and social studies to devote more time to tested math skills (Salmon-Cox 1982, 1984). Darling-Hammond and Wise (1985) also found that tested areas were taught at the expense of untested areas; McNeil (1988) documented that minimum proficiencies dictated by competency tests overwhelmed class time.

In addition, Darling-Hammond and Wise (1985) found that even within the bounds of test-driven content there was "dumbing down" of instruction. Teachers taught the precise content of the tests rather than underlying concepts; and skills were taught in the same format as the test rather than as they would be used in the real world. For example, teachers reported giving up essay tests because they are inefficient in preparing students for multiple-choice tests.

Conceiving instruction in the format of multiple-choice items has other far-reaching negative consequences: it leads to endless drill and practice on decontextualized skills. The notion that learning comes about by the accretion of little bits is outmoded learning theory. Current models of learning based on cognitive psychology contend that learners gain understanding when they construct their own knowledge and develop their own cognitive maps of the interconnections among concepts and facts. Thus, real learning cannot be spoon-fed, one skill at a time.

Logically, then, to become adept at thinking and reasoning, students need practice in solving real problems and comprehending complex texts. Not surprisingly, students given instruction aimed at conceptual understanding do better on skills tests than students drilled on the skills directly (Carpenter et al. 1988). Thus, the practice of postponing higher-order thinking goals until low-level skills have been mastered is harmful. Low-achieving students suffer most from a proficiency-driven curriculum because they are consigned indefinitely to dull and repetitive skills instruction that does not enable them to grasp underlying concepts (Levin 1987).

Students given instruction aimed at conceptual understanding do better on skills tests than students drilled on the skills directly.

Moreover, test-defined instruction has the effect of driving out good teachers and "deskilling" those who remain (McNeil 1988). Teachers with the least content knowledge might feel secure in leading students through worksheets fashioned after the tests, even if they become less skilled as teachers in the process. Teachers who know the most about content, however, are insulted when

external mandates prevent them from using their own expertise to devise instruction. They must either find a way to resist the deskilling, as in McNeil's magnet schools, capitulate, or get out of teaching.

Finally, teaching to the test devalues the meaning of the test results themselves. By having students practice on "remarkably similar" items, teachers can improve their test performance, but gains from this type of teaching do not necessarily generalize to independent measures of the same content. Thus, test scores can go up without a commensurate gain in achievement. To really assess what students know, legislators and school boards would need a completely new test for each administration.

SUBSTANTIVELY BETTER ASSESSMENTS

Twenty years ago, standardized tests served as reasonable indicators of student learning. In today's political climate, tests are inadequate and misleading as measures of achievement. Assessment tasks should be redesigned—indeed, are being redesigned—to more closely resemble real learning tasks. Tests should require more complex and challenging mental processes from students. They should acknowledge more than one approach or one right answer and should place more emphasis on uncoached explanations and real student products. Structured formats should be changed often enough that there can be no benefit to practicing a skill in one particular format. Further, the dimensions of the test domain must be expanded so that teaching to the test does not imply teaching only a subset of learning goals.

At a recent conference on Assessment in the Service of Learning, Robert Glaser (1988) identified several indicators of competence in a domain of knowledge. Assessment should collect these types of evidence:

- *Coherence of knowledge.* Beginners' knowledge is spotty and superficial, but as learning progresses, understanding becomes integrated and structured. Thus assessment should tap the connectedness of concepts and the student's ability to access "interrelated chunks."

- *Principled problem solving.* Advanced learners ignore the surface features of a task and recognize underlying principles and patterns needed to solve the problem.
- *Knowledge use.* Complete understanding also means knowing the conditions that mediate the use of knowledge.
- *Automatized skills.* Basic component skills must be automatized so as to be integrated into total performance. (This is the only indicator that resembles today's by-rote measures of skills.)
- *Metacognitive or self-regulatory skills.* Assessment should determine whether students are able to monitor their own understanding, use strategies to make questions comprehensible, evaluate the relevance of accessible knowledge, and verify their own solutions.

The best way to check for these indicators is to make assessment measures resemble learning tasks. In current practice, some of the best examples of more learning-like tests are writing assessments. So long as language mechanics are not overemphasized and ideas and coherence of expression submerged, the test product represents important learning; and practice for the test is practice with real writing. Another way to assess achievement is by collecting portfolios of students' work. In the Vermont assessment proposal, for example, students and teachers would select samples of each student's best written work in English, social studies, and science. Besides encouraging more thoughtful activities in the classroom, such an approach can "enhance teacher professionalism by enabling classroom teachers to get involved in developing and scoring the assessment" (Rothman 1988).

The dimensions of the test domain must be expanded so that teaching to the test does not imply teaching only a subset of learning goals.

CLASSROOM ASSESSMENT

Better content is not the only answer to improved assessment. Policymakers must also understand and preserve

the important distinction between classroom assessment and accountability testing. Cole (1988) explained the incompatibilities between measurement to serve accountability and policy goals and measurement for instructional purposes. Large-scale assessments must be formal, objective, time-efficient, cost-efficient, widely applicable, and centrally processed. Most important, results must be in a form useful to policymakers, which usually means reducing complexity to a single score. In contrast, assessments designed to support instruction are informal, teacher-mandated, adapted to local context, locally scored, sensitive to short-term change in students' knowledge, and meaningful to students. They provide immediate, detailed, and complex feedback; and they incorporate tasks that have instructional value in themselves.

Classroom assessment is conducted in a climate of greater trust than are standardized tests. Classroom tests and observations do not have to meet the same standards of accuracy. Errors made in judging individual students are less serious and more easily redressed as teachers gather new evidence. Although single teacher tests are probably less reliable (in a statistical sense) than a one-hour standardized test, the accumulation of data gathered about individual pupils in the course of a school year has much more accuracy.

Classroom assessment places no political pressure on teachers; it can address the full range of learning goals. The low-pressure environment allows the teacher to ask potentially ambiguous questions that elicit higher-order thinking. In classroom tests, teachers can continue to reach for the conceptually more important questions because they have the luxury of responding to student questions during the assessment, allowing them to restructure the problem as they understand it, and accepting more than one answer as correct. In contrast, accountability tests cannot be experimental in nature; questions aimed at more ambitious constructs are left out if they might not withstand legal scrutiny.

The differences between accountability and instructional assessment are so fundamental and necessary that it may not be desirable to merge the two purposes. Instead,

efforts to pursue an agenda like Glaser's, to develop formal measures of integrated understandings, and to train teachers in informal methods will have greater integrity if such procedures are *not* turned into accountability devices. While the substance of accountability tests must be improved by keeping conceptions of real learning in mind, their susceptibility to distortion should not be allowed to contaminate classroom uses of student data.

SAFEGUARDING ACCOUNTABILITY ASSESSMENT

To restore the credibility of accountability tests, we must remove both the incentives and the means to distort scores. For example, rewarding high test scores with bonuses to schools or merit pay to teachers is a clear invitation to teach to the test. Because teaching to the test can raise scores more dramatically than can instruction designed to improve achievement (Shepard 1988), the incentive system could reward the worst practices.

The more we focus on raising test scores, the more instruction is distorted, and the less credible are the scores themselves.

We must institute procedural safeguards to signify that scores will not be used punitively and thereby protect the meaning of the test data. For example, the use of scientific sampling to report for an entire state reduces the threat to individual classrooms, because classroom results are not reported. Similarly, a fall testing date removes the ownership for scores from the teachers who supervise test administration, thereby reducing the incentive to redirect instruction to the test. When sampling is used, policymakers relinquish the ability to rank every teacher in the state by test scores but gain believable data. If school and district results are announced to the media, data should also be reported on the wealth of each community and the skills of incoming students, to avoid bragging by rich districts and punishment of poor districts.

The most important safeguard is not to give the same test repeatedly, year after year. Although using alternate forms of a minimum skills test does not prevent the teaching of too narrow a set of goals, new forms at least

preclude rehearsing the specific vocabulary or word problems found on any one form of a test.

Another procedure designed to protect breadth of content is matrix sampling, used by the California Assessment Program. Every pupil at a grade level takes a test that is in fact a randomly assigned subpart of a larger test. Scores cannot be computed for individual pupils. However, the resulting scores reported for schools, districts, and the State of California are based on a broad content domain, making it less likely that teaching to the test will limit instruction to a narrow set of skills. Carrying this idea further, large-scale assessments could cover many broad domains by using a cycle of topics so that different subject areas could be comprehensively assessed each year.

Ironically, the National Assessment of Educational Progress began almost 20 years ago with many of these substantive and procedural features, including the assessment of 10 subject areas in 4-year cycles. These features have gradually been altered in response to cost constraints and the shift from a monitoring purpose to accountability. The proportion of performance tasks compared to paper-and-pencil measures has decreased. The number of subject areas has been reduced in favor of more frequent assessments in the basics. Originally, data were reported at the item level, of use only to curriculum experts. Today results are aggregated into a single score for each subject area to produce a national report card. In 1968 a sampling design based on states was purposely avoided to prevent invidious comparisons and to ensure local cooperation with the assessment. In keeping with today's climate, however, state-by-state comparisons have been mandated by Congress beginning with the 8th grade mathematics assessment of 1990.

EXPANDING BOUNDARIES

Accountability testing in the 1980s is having a pernicious effect on education. Standardized tests have always been fallible, limited measures of learning goals. Today, the joint effect of limited content and inordinate weight attached to test results is magnifying the drawbacks. Tests are being asked to do too much. The more we focus on

raising test scores, the more instruction is distorted, and the less credible are the scores themselves.

The content boundaries of current assessments should be expanded to include conceptual understanding and problem-solving abilities. However, substantive improvements alone will not correct current ills. If policymakers want assessment data to reflect accurately what students know, they must remove incentives to teach to the test.

REFERENCES

Cannell, J. J. (1987). *National Normed Elementary Achievement Testing in America's Public Schools: How all Fifty States are Above the National Average.* Daniels, W. V.: Friends for Education.

Carpenter, T. P, E. Fennema., P. L. Peterson., C. Chiang, and M. Loef. (April 1988). "Using Knowledge of Children's Mathematical Thinking in Classroom Teaching: An Experimental Study." New Orleans: Paper presented at the annual meeting of the American Educational Research Association.

Cole, N. S. (1988). "A Realist's Appraisal of the Prospects of Unifying Instruction and Assessment." In *Assessment in the Service of Learning: Proceedings of the 1987 ETS Invitational Conference.* Princeton, N.J.: Educational Testing Service.

Darling-Hammond, L., and A. E. Wise. (1985). "Beyond Standardization: State Standards and School Improvement." *The Elementary School Journal,* 85: 315-336.

Glaser, R. (1988). "Cognitive and Environmental Perspectives on Assessing Achievement." In *Assessment in the Service of Learning: Proceedings of the 1987 ETS Invitational Conference.* Princeton, N.J.: Educational Testing Service.

Levin, H. M. (March 1987). "Accelerated Schools for Disadvantaged Students." *Educational Leadership* 44: 19-21.

McNeil, L. M. (1988). "Contradictions of Control, Part 3: Contradictions of Reform." *Phi Delta Kappan* 69: 478-485.

Popham, W. J. (1987). "The Merits of Measurement-Driven Instruction." *Phi Delta Kappan* 68: 679-682.

Rothman, R. (October 26, 1988). "Vermont Plans to Pioneer with 'Work Portfolios.'" *Education Week* 8, 1: 11.

Salmon-Cox, L. (September 1984). "MAP Math: End of Year One Report." Pittsburgh: Learning Research and Development Center.

Salmon-Cox, L. (September 1984). "MAP Reading End of Year Report." Pittsburgh: Learning Research and Development Center.

BEST COPY AVAILABLE

Shepard, L. A. (April 1988). "Should Instruction Be Measurement-Driven?: A Debate." New Orleans: Paper presented to the annual meeting of the American Educational Research Association.

Tyson-Bernstein, H. (1988). "America's Textbook Fiasco: A Conspiracy of Good Intentions." *American Educator* 12: 20-27, 39.

Testing and Thoughtfulness

The new literacy of thoughtfulness calls for a quite different technology of teaching and testing; it rests on quite different notions about the nature of knowledge and learning because it is about the making of meaning, not just the receiving of it. Thoughtfulness is a constructive, not a passive, undertaking.

–Rexford Brown

In order for testing to become a more powerful tool of education and an effective way to measure student growth, it must move from the end-of-the-week, end-of-the-chapter, end-of-the-unit summative form of final evaluation to ongoing formative assessment. The purpose of testing should be to monitor student growth and to measure thoughtfulness during the process of learning. Our emphasis on the final product—the test results—represents our over-emphasis on facts, content, and the curriculum, often at the expense of higher-order thinking skills, problem solving, diverse learning styles, reflectiveness, and meaningful interactions.

If we are to measure student growth, we need to shift from traditional paper-and-pencil, fill-in-the-blank, short-term memory scantron tests to authentic assessment tasks where students can demonstrate what they know or do not know by using a wide variety of formats. Students engaged in active learning can become lifetime learners. If we treat them passively as "empty vessels," to store discrete bits of factual information, we impede their ability to see the relevance of the material. We also impede their ability to make connections between the material memorized for the test and the application of that material to real life.

We need to design authentic tasks that require students to use higher literacies. Rexford Brown says "These higher literacies call for students to analyze, think critically, evaluate, synthesize information, communicate more effectively, solve problems, learn how to learn, and, in general, learn far more actively than traditionally."

Grant Wiggins recommends that we "test those capacities and habits we think are essential, and test them in context. Make them replicate, within reason, the challenges at the heart of each academic discipline. Let them be—authentic."

Instead of building a curriculum around discrete bits of knowledge, each faculty should decide on significant goals or outcomes they want their students to achieve for the school year and beyond. If these goals are important, if they are goals worth teaching, then they are worth testing. If we concentrate on making students memorize facts that they will forget within a week and that are not meaningful to them, most of them will give the facts back to us on the test. Shortly after the test those facts will probably escape the brain forever — without a trace! If, however, the learning is significant and the students have internalized the concepts through interactions with other students, then the chance that the information will be stored in the short-term or long-term memory is much greater. As Elliot Eisner noted, "If it is not worth teaching, it is not worth testing!"

Wiggins feels that authentic tests have four basic characteristics in common:

- They are designed to be truly representative of performance in the field.
- Far greater attention is paid to the teaching and learning of the *criteria* to be used in the assessment.
- Self-assessment plays a much greater role than in conventional testing.
- Students are often expected to present their work and defend themselves publicly and orally to ensure that their apparent mastery is genuine.

In addition to teaching for the authentic test, many teachers are conducting ongoing assessment that allows for individual learning styles. Project Spectrum was

designed to recognize variation in learning styles by identifying young children's distinctive cognitive and stylistic profiles.

Mara Krechevsky says that "Spectrum is based on the assumption that every child has the potential to develop strength in one or several content areas and that it is the responsibility of the educational system to discover and nurture these proclivities." Every attempt must be made to address learning styles and allow students to perform using a variety of the multiple intelligences.

Testing can be an effective tool to measure student growth. It is essential, however, that we construct tests that challenge students to think critically and creatively rather than to memorize and recall information. The new "literacy of thoughtfulness" requires students to become actively involved in the learning process and to apply these skills not just to the semester exam but to real life. As Rexford Brown states, "Students who have to perform or exhibit their knowledge and skills get learning in their bones: active learners become lifetime learners."

Testing and Thoughtfulness

by Rexford Brown

The literacy of thoughtfulness calls for a new concept of testing, one that reflects the active nature of learning.

The definition of literacy that sufficed for an earlier generation and a different economy has been replaced by a host of higher literacies: computer, scientific, civic, cultural, and so on. These higher literacies call for students to analyze, think critically, evaluate, synthesize information, communicate more effectively, solve problems, learn how to learn, and, in general, learn far more actively than traditionally. They call for greater thoughtfulness, both in the thinking sense and in the sense of caring more about other people and one's community.

Higher literacies call for students to analyze, think critically, evaluate, synthesize information, communicate more effectively, solve problems, learn how to learn, and, in general, learn far more actively than traditionally.

For the past two years, with a grant from the John D. and Catherine T. MacArthur Foundation, we at the Education Commission of the States have been investigating the perceived and real effects of policy upon efforts to develop a much higher level of literacy for a much broader range of students. This "Policy and the Higher Literacies" project came about because education reformers, politicians, and business leaders are calling for larger numbers of graduates to be good thinkers, problem solvers, and risk takers.

POLICY AND THE HIGHER LITERACIES

What, we have been asking ourselves, are state and local education policies doing to promote greater thoughtful-

From *Educational Leadership*, vol. 46, no. 7, p. 31-33, 1989. Reprinted with permission.

ness in school or to inhibit it? We have been looking carefully at policies in curriculum (including textbook selection), teacher certification and professional training, school finance, and testing and accountability. We have interviewed experts, held focus group meetings, reviewed all available research, and conducted case studies of districts in six regions.

In our case studies, we reviewed the state and local policies and guidelines affecting thoughtfulness (e.g., mandates mentioning thinking or problem solving); interviewed appropriate state and local administrators; and then visited schools and classrooms (grades 3, 6, 8, and 11) where "higher literacy" activities were purportedly going on. All told, in these case studies we gathered 650 hours' worth of interview and observational information, about two-thirds of which was derived from talking with teachers and observing their classes.[1]

Almost everyone we talked to is determining educational success or progress on the basis of scores on commercial standardized norm-referenced tests. This, of course, is not news to anyone who has been watching American education over the last 20 years. Schools, principals, and teachers differ markedly, however, in their perceptions about this fact and the significance they attach to it. What it means to use these tests depends on who you are and what kind of pressure you feel you are under. By and large, schools that are not in the public doghouse or under tight court monitoring do their testing routinely, make little use of the results, and go about their business. On the other hand, schools under public scrutiny (even for reasons having nothing to do with achievement) encourage much more teaching to the test and make much more out of declines or increases in scores.

By and large, schools that are not in the public doghouse or under tight court monitoring do their testing routinely, make little use of the results, and go about their business.

Analogously, experienced, confident teachers pay little attention to externally imposed tests. They don't need them, except, perhaps, to reinforce judgments they have already made on the basis of their own more comprehen-

sive information about students. If test results contradict what they know about students, they tend to ignore them. New teachers, however, and underprepared or insecure or embattled teachers, are more likely to teach to the tests, even if they believe the tests are poor indicators of student learning.

Next, most people we talked to believe that commercial standardized tests are valid indicators of student and school achievement. Although many researchers and even testing experts see serious limitations to these tests and caution users constantly about misuse or misinterpretation, school people seldom showed awareness or concern about these matters. When we asked whether the tests adequately addressed higher-order thinking skills or performances (they do not), they either didn't know, didn't care, or believed the test manufacturer had dealt with that problem.

Third, most people we talked to about thinking, creativity, and problem-solving programs believed that these kinds of activities cannot be tested at all or cannot be tested objectively. They would invest in more innovative programs, they said, if inexpensive tests could be developed for them.

We also found that in certain large districts thinking skills are being taught by breaking thinking itself into a collection of facts, terms, and subactivities that can be taught and tested one at a time. Our observations of these approaches led us to conclude that whatever else was going on, students were *not* engaged in active thinking or problem solving or creativity. Thoughtfulness itself had been adapted to the prevailing basic skills transmission model of teaching and testing.

What we found, in short, was considerable passivity with respect to testing, mixed with lackadaisical use of results and a dependency that, under certain circumstances, leads to shaping of teaching and curriculum around the tests. People engaged in higher literacy activities did not worry much about the standardized tests, correctly assuming their students would do fine on them. People who were looking for excuses not to undertake higher literacy activities complained that they could not be

tested at all or at least not within the current accountability framework. Educators trying nominally to introduce critical and creative thinking and problem-solving activities had reshaped those activities so they would fit the prevailing technology.

THE ACTIVE NATURE OF LEARNING

We did see thoughtfulness in action in several places. Ironically—and instructively—the most exciting programs were in a district that uses *no* standardized tests or assessments. No one there felt much need for them, either to prod the schools toward higher achievement or to measure current achievement. Yet the schools were among the very best we visited, and the students were among the most energized we met. We did not doubt that they were learning constantly and well, regardless of their backgrounds, which were as diverse as you can find.

The primary rub between thoughtfulness and the kinds of testing to which we are addicted is the rub between active and passive learning. The new literacy of thoughtfulness calls for a quite different technology of teaching and testing; it rests on quite different notions about the nature of knowledge and learning because it is about the making of meaning, not just the receiving of it. Thoughtfulness is a constructive, not a passive, undertaking.

The prevailing teaching and testing technology rests on the assumption that knowledge is objective and can be drilled into passive "blank-slate" brains, then paraded out on cue. This may apply to some kinds of knowledge—for instance, multiplication tables. But most knowledge is socially and personally derived; we are creating it all the time through social interaction, and its nature and uses constantly shift. Moreover, adults spend most of their time sifting through and using various "knowledges" rather than "arriving at" them. They have to analyze, synthesize, interpret, and evaluate facts and ideas far more often than

The prevailing teaching and testing technology rests on the assumption that knowledge is objective and can be drilled into passive "blank-slate" brains, then paraded out on cue.

they have to "know" them in the sense of only being able to recite them. These are precisely the kinds of thoughtful abilities that more and more leaders say graduates should possess and that fewer and fewer graduates actually do possess. And they are precisely the activities that cannot be tested the way we currently test students in most schools.

A NEW CONCEPT OF TESTING

What needs to change in order to make room for more thoughtfulness in the schools is not just the tests—which clearly do not require students to create, construct, negotiate, and communicate meaning—but the whole concept of testing itself. The concept that testing is initiated externally from the student, separate from the learning process, and primarily aimed at determining whether inert knowledge is in students' short-term memories exercises far too much influence over school people today. The goals of thoughtfulness are that students internalize capacities to evaluate their learning, do so as they learn, and do so in ways that exhibit their capacity to be *performing* thinkers, problem solvers, and inquirers. The dominant technology neither tells us whether those goals are being met, nor encourages anyone to find out, nor models a kind of inquiry into achievement that students and teachers might profitably imitate.

Students who have to perform or exhibit their knowledge and skills get learning in their bones: active learners become lifetime learners.

The best tests of thoughtfulness I saw in my travels were student generated. Nothing helps you master a subject better than having to ask and debate fundamental questions about what is most important about the subject and how you could tell if someone else had mastered it. After all, higher literacy is acquired more through a "do as I do" than a "do as I say" relationship between teachers and learners. It is developed through experience more than through lecture, through practice more than through passivity.

Students of all ages who create some of their own examinations are forced to reflect on what they have studied and make judgments about it. They are forced to

ask questions and challenge answers. They come face to face with the fact that 30 students derive 30 different interpretations from the same lesson, and they have to evaluate and reduce those differences. They learn how to ask the right questions and how to know good answers from bad, irrelevant from relevant. Students who have to perform or exhibit their knowledge and skills get learning in their bones: active learners become lifetime learners.

It may appear that if too many schools turned to open-ended testing, portfolios, performances, and exhibitions, no one would be able to compare their effectiveness, and policy development would suffer as a result. This need not become a problem. There is no reason to do away entirely with standardized tests that do not promote cheating (e.g., everyone's being "above average"). The task is to keep them from dominating school affairs to such a degree that they distort curriculum and hold back efforts to make schools more thoughtful places. Moreover, professionals in other fields have found ways to aggregate "messy" information systems for policy purposes. This difficult but not impossible enterprise should have high priority in state departments of education and educational labs and centers.

SCHOOLS AS THOUGHTFUL ENVIRONMENTS

What you as a professional can do right now is to scrutinize tests and testing practices and determine to what degree they constrain efforts to teach critical and creative thinking, problem solving, the reasoning arts, and other dimensions of thoughtfulness in your school. If you find evidence of constraints, start a sustained schoolwide inquiry into ways you can remove them and demonstrate learning in more concrete, robust, and exciting ways than filling in ovals. You may find that you will have begun restructuring your school without even trying. At the very least, you will create a thoughtful environment that will model the very behavior you most want to cultivate in your students.

NOTES

1. Our report, *Schools of Thought*, is in preparation.

Project Spectrum: An Innovative Assessment Alternative

by Mara Krechevsky

By evaluating young children's strengths in many domains, not just language and logic, the Spectrum battery promises to give all children a chance to shine.

It's free choice time in the afternoon 4-year-olds' classroom at the Eliot-Pearson Children's School in Medford, Massachusetts. The class is participating in Project Spectrum, an innovative approach to assessment in early childhood. Hallie, not quite 4 years of age, is once again roaming from activity to activity, finding it difficult to concentrate on a task for longer than five minutes at a time. Her ever-patient teacher first tries engaging her in the art table project, then in experimenting with the siphons at the water table, and finally in playing hospital at the dramatic play area, all to little avail. Hallie is easily distracted; she becomes sillier as the hour wears on, making increasingly poor eye contact with the teacher and speaking in nonsense words.

Spectrum is based on the assumption that every child has the potential to develop strength in one or several content areas.

With 20 minutes left until group time, Hallie's turn comes up for the week's Spectrum activity: the assembly task. This task involves taking apart and reassembling two food grinders. Successful completion of the activity depends on a combination of fine motor skills with visual-spatial and problem-solving abilities. Hallie's eyes light up expectantly upon seeing the first grinder, and she immediately touches the main fastener, which loosens the handle and inner mechanism. Within minutes, the grinder is completely disassembled, and Hallie begins to put the

From *Educational Leadership*, vol. 48, no. 5, p. 43-48, 1991. Reprinted with permission.

pieces back together, carefully figuring out the correct direction in which to turn each screw. She adopts a trial-and-error approach, her feet swinging excitedly up and down from her chair whenever she succeeds. She remains focused, persistent, and methodical throughout—correcting her own mistakes, oblivious to the rest of the class.

This anecdote illustrates the power of the Spectrum approach to assessment. Spectrum began in 1984 at Harvard and Tufts Universities as an attempt to reconceptualize the traditional linguistic and logical/mathematical bases of intelligence. Our first four years of research centered on identifying young children's distinctive cognitive and stylistic profiles. In addition to assessing linguistic and mathematical abilities, the Spectrum assessment battery examines mechanical, spatial, bodily, musical, social, and scientific abilities as well (see Fig. 1). In a political climate that places increasing pressure on educators to extend formal instruction downward and embrace a narrow view of scholastic readiness, Spectrum offers a developmentally appropriate alternative based on a broad view of the mind.

A RICH CLASSROOM ENVIRONMENT

The theoretical foundation of the project stems from Howard Gardner's (1983) theory of multiple intelligences and David Feldman's (1980) theory of development in non-universal domains. Although many early childhood educators still think in terms of children's progressing through broad, undifferentiated stages of universal development, Spectrum was designed to recognize variation in both children and areas of activity. Thus, the Spectrum model identifies domain-specific strengths in areas often not included in many Piagetian or neo-Piagetian approaches to education. Spectrum is based on the assumption that every child has the potential to develop strength in one or several content areas and that it is the responsibility of the educational system to discover and nurture these proclivities. Rather than building around a test, the Spectrum approach is centered on a wide range of rich activities; assessment comes about as part-and-parcel of the child's involvement over time in these activities.

As Figure 1 indicates, the Spectrum measures range from relatively structured and targeted tasks (for example, in the number and music domains) to less structured measures and observations (for example, in the science and social domains). These measures form one part of a rich classroom environment that is equipped with engaging materials, games, puzzles, and learning areas. The learning areas enable children to make initial explorations of materials related to the domains assessed by Spectrum, as well as offering follow-up activities. For example, after the storytelling task, children can be asked to create their own storyboards at the art area. The assessment activities are administered throughout the school year. Documentation takes a variety of forms, from score sheets and observation checklists to portfolios and tape recordings.

A LOOK AT SPECTRUM'S FACETS

Distinctive features of the Spectrum assessment system include:

1. **Blurring the line between curriculum and assessment.** By gathering information over time in the child's own environment, Spectrum effectively blurs the traditional division between curriculum and assessment. For example, teachers collect children's artwork in portfolios and observe bodily-kinesthetic abilities through a biweekly creative movement session. The traditional test setting of a small room with an unfamiliar examiner administering timed and standardized instruments, in the Spectrum view, provides too narrow and skewed a view of the child. In the Spectrum tasks, children's skills are integrated, rather than isolated. Thus, as we saw earlier, the assembly activity engages Hallie in an applied and meaningful task presented as part of her preschool curriculum.

2. **Embedding assessment in meaningful, real-world activities.** Rather than just focusing on skills useful in the school context, Spectrum uses the concept of adult *endstates* to focus its assessments on abilities relevant to achieving significant and rewarding adult roles. Examples

Figure 1: Areas of Cognitive Ability Examined in Project Spectrum

NUMBERS

Dinosaur Game: Measures a child's understanding of number concepts, counting skills, ability to adhere to rules, and use of strategy.

Bus Game: Assesses a child's ability to create a useful notation system, perform mental calculations, and organize number information for one or more variables.

SCIENCE

Assembly Activity: Measures a child's mechanical ability. Successful completion of the activity depends on fine motor skills and visual-spatial, observational, and problem-solving abilities.

Treasure Hunt Game: Assesses a child's ability to make logical inferences. The child is asked to organize information to discover the rule governing the placement of various treasures.

Water Activity: Assesses a child's ability to generate hypotheses based on his or her observations and to conduct simple experiments.

Discovery Area: Includes year-round activities that elicit a child's observations, appreciation, and understanding of natural phenomena.

MUSIC

Music Production Activity: Measures a child's ability to maintain accurate pitch and rhythm while singing and his or her ability to recall a song's musical properties.

Music Perception Activity: Assesses a child's ability to discriminate pitch. The activity consists of song recognition, error recognition, and pitch discrimination.

LANGUAGE

Storyboard Activity: Measures a range of language skills including complexity of vocabulary and sentence structure, use of connectors, use of descriptive language and dialogue, and ability to pursue a storyline.

Reporting Activity: Assesses a child's ability to describe an event he or she has experienced with regard to the following criteria: ability to report content accurately, level of detail, sentence structure, and vocabulary.

VISUAL ARTS

Art Portfolios: The contents of a child's art portfolio are reviewed twice a year and assessed on criteria that include use of lines and shapes, color, space, detail, and representation and design. Children also participate in three structured drawing activities. The drawings are assessed on criteria similar to those used in the portfolio assessment.

MOVEMENT

Creative Movement: The ongoing movement curriculum focuses on children's abilities in five areas of dance and creative movement: sensitivity to rhythm, expressiveness, body control, generation of movement ideas, and responsiveness to music.

Athletic Movement: An obstacle course focuses on the types of skills found in many different sports such as coordination, timing, balance, and power.

SOCIAL

Classroom Model Activity: Assesses a child's ability to observe and analyze social events and experiences in his or her classroom.

Peer Interaction Checklist: A behavioral checklist is used to assess the behaviors in which children engage when interacting with peers. Different patterns of behavior yield distinctive social roles such as facilitator and leader.

of endstates include *naturalist, salesperson, singer, dancer,* and *social worker.* Thus, in the language domain, Spectrum examines a child's ability to tell stories or provide a descriptive account of an experience—valuable skills for novelists and journalists—rather than his or her ability to repeat a series of sentences. For Hallie, the applicable endstate is *mechanic.* In contrast to many standardized assessments, which might have Hallie copy shapes or block patterns, Spectrum provides her with a real machine to work on. This grounding of assessments in real-world activities ensures that the areas addressed are likely to be meaningful to the child, the teacher, and the child's family.

Spectrum's grounding of assessments in real-world activities ensures that the areas addressed are likely to be meaningful to the child, the teacher, and the child's family.

3. Using measures that are "intelligence-fair."

Rather than viewing all abilities through the window of language and logic, as most standardized tests do, Spectrum attempts to tap abilities directly, via their own particular medium. In the above anecdote, Hallie works directly with simple mechanical objects, rather than answering questions about how machines work. The music perception and production tasks of the assessment employ Montessori bells and simple songs, while the movement activities elicit both athletic and creative movement.

4. Emphasizing children's strengths.

In contrast to many educational approaches, particularly those used with children at risk for school failure, the Spectrum assessment approach seeks to identify children's areas of strength and to construct their education as much as possible around those domains of competence. Giving children experience in their areas of strength might not only increase their sense of self-esteem but suggest ways to address areas that are not as strong. For example, to boost Hallie's language skills, she could be asked to dictate instructions for disassembling a grinder or to tell a story about a machine she might invent.

5. Attending to the stylistic dimensions of performance.

In order to capture fully a child's approach to a task, we soon discovered it was important to look not only at a child's cognitive skills but at certain stylistic features as well. "Working styles" describe how a child interacts with the materials of a domain, such as his or her persistence, attention to detail, and level of confidence (see Fig. 2). While some children exhibit the same working style across domains, others have styles that are much more content-specific. Such information has important implications for designing educational interventions. In Hallie's case, she revealed the capacity to become extremely focused and reflective when working in her area of strength.

Figure 2: Stylistic Features Examined in Project Spectrum

CHILD IS:

- easily engaged/reluctant to engage in activity
- confident/tentative
- playful/serious
- focused/distractible
- persistent/frustated by task
- reflective about own work/impulsive
- apt to work slowly/apt to work quickly

CHILD:

- responds to visual/auditory/kinesthetic cues
- demonstrates planful approach
- brings personal agenda/strength to task
- finds humor in content area
- uses materials in unexpected ways
- shows pride in accomplishment
- shows attention to detail/is observant
- is curious about materials
- shows concern over "correct" answer
- focuses on interaction with adult
- transforms task/material

In the Spectrum approach, all of the information collected on a child is compiled at year's end into a "Spectrum Profile": a short description, written in nontechnical prose, of the child's participation in the project's activities. The report addresses each child's areas of strength, either relative to himself or herself or to the child's peer group. Because of our belief that psychologists spend too much time ranking children and not enough time trying to help them, we also give concrete suggestions for follow-up activities that can be carried out at home or in the community. A "Parent Activities Handbook" suggests home

activities that use inexpensive and easily acquired materials (for example, different ways to grow seeds, measuring and counting games, and so on).

EARLY RESEARCH RESULTS

Hallie's class was one of two Eliot-Pearson preschool classrooms that participated in the initial phase of Spectrum research. This phase focused on assessing individual children's cognitive and stylistic strengths. Preliminary results from the two classrooms suggest that the Spectrum system does indeed identify distinctive intellectual profiles in young children[1] (for a full report, see Gardner and Hatch 1989, Krechevsky and Gardner 1990). We also found some evidence that a child's strength in one area might facilitate performance in another. For example, one child demonstrated exceptional storytelling ability, yet generally refused to participate in creative movement. However, she moved with unusual expressiveness when presented with storyboard props as a stimulus during one of the movement sessions.

With regard to working styles, most children seemed to exhibit domain-specific configurations. Many were reflective and attentive to detail only in their areas of strength. However, some children demonstrated a more general working style, which at times worked to their disadvantage. For example, one boy approached every activity with his own agenda of ideas. Although in the less structured environment of the class he conducted many compelling experiments to test his hypotheses, he was unable to adjust to more structured task situations.

Preliminary follow-up data on children in the original Spectrum class indicate that strengths and working styles remained roughly constant one to two years later. Sometimes the particular combination of a child's areas of strength and working style determined whether or not a strength would re-emerge. One girl, who in preschool constantly sought the positive regard of her peers, spent a lot of time at the writing table because her language abilities were advanced for her age group. However, because she was not the most able in her group the following year, she devoted the majority of her free time to

art activities. Thus, the language abilities identified earlier were less likely to resurface and develop.

Responses from parents indicated that the areas where they were most surprised to learn of strengths included music perception, mechanical ability, and creative movement. A number of parents in the follow-up found it very useful to have a written profile to which they could compare more recent views of their child. The area parents were most likely to encourage at the one-year follow-up was drama, perhaps because they saw it as an especially effective way to combine ability in the language, social, and movement domains.

Currently, the Spectrum approach is being modified for use with children in kindergarten and 1st grade and for children who are more at risk for school failure. A broad-based Spectrum curriculum is being implemented in selected 1st grade classrooms with a large at-risk population in Somerville, Massachusetts. Children are being encouraged to develop their areas of strength in an apprentice-type model and to bring these strengths to bear on the established curricular goals of the 1st grade. We are administering pre- and post-tests to determine change in children's academic achievement, self-esteem, attitude toward school, and school adjustment.

Children are encouraged to develop their areas of strength and to bring these strengths to bear on established curricular goals.

THEIR PLACE IN THE SUN

It may be important at this point to outline the disadvantages and the advantages of Spectrum. First, a pluralized model of intelligence runs the risk that achievement-oriented parents will push their children in 15 areas, instead of a few. Second, parents and teachers may be tempted to track a child prematurely into pursuing his or her area(s) of strength. Finally, parents outside the mainstream culture may not be as concerned with their children's performance in domains not valued by the traditional culture.

Nevertheless, the approach offers a number of benefits. The Spectrum battery exposes children to more domains than are typically included in early childhood assessments or curriculums. Spectrum also actively involves children in the assessment process. They collect their work for the art portfolios and tape their own stories and songs. Time is also set aside for children to reflect on the activities in general. The Spectrum approach can be used on many levels: as assessment, as curriculum, or as a powerful philosophical framework through which to view children and their particular sets of strengths and working styles. In fact, the approach is as much a framework and set of ideas as it is a discrete program. Spectrum is currently being adapted for a variety of purposes by both researchers and practitioners in the field.[2] Because the approach takes individual differences seriously, it enables teachers to accommodate diverse populations and to individualize their curriculums. Moreover, because of its provision of many ways to demonstrate excellence, including ways that go beyond conventional scholastic success, Spectrum may be particularly suited for at-risk children. At best, the Spectrum approach promises to increase the chances for all children to find their place in the sun.

Author's note: The work described in this article was supported by generous grants from the William T. Grant Foundation, the Rockefeller Brothers Fund, and the Spencer Foundation.

NOTES

1. Because of the small sample size, the results reported in this article should be regarded as tentative.
2. We encourage these efforts and are interested in hearing from people who have tried to implement Spectrum. However, for reasons of logistics and limited resources, at present we cannot provide much in the way of support.

REFERENCES

Feldman, D. H. (1980). *Beyond Universals in Cognitive Development.* Norwood, N.J.: Ablex.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences.* New York: Basic Books.

Gardner, H., and T. Hatch. (November 1989). "Multiple Intelligences Go to School: Educational Implications of the Theory of Multiple Intelligences." *Educational Researcher 18*, 8: 4-10.

Krechevsky, M., and H. Gardner. (1990). "The Emergence and Nurturance of Multiple Intelligences: The Project Spectrum Approach." In *Encouraging the Development of Exceptional Skills and Talents*, edited by M.J.A. Howe. Leicester, U.K.: The British Psychological Society.

Teaching to the (Authentic) Test

by Grant Wiggins

Testing can once again serve teaching and learning if tests clarify and set intellectual standards.

Practical alternatives and sound arguments now exist to make testing once again serve teaching and learning. Ironically, we *should* "teach to the test." The catch is to design and then teach to *standard-setting* tests so that practicing for and taking the tests actually enhances rather than impedes education, and so that a criterion-referenced diploma makes externally mandated tests unobtrusive—even unnecessary.

SETTING STANDARDS

If tests determine what teachers actually teach and what students will study for—and they do—then the road to reform is a straight but steep one: test those capacities and habits we think are essential, and test them in context. Make them replicate, within reason, the challenges at the heart of each academic discipline. Let them be—authentic.

If tests determine what teachers actually teach and what students will study for—and they do—then test those capacities and habits we think are essential, and test them in context.

What are the actual performances that we want students to be good at, that represent model challenges? Design them by department, by school, and by district—and worry about a fair, efficient, and objective method of grading them as a *secondary* problem. Do we judge our students to be deficient in writing, speaking, listening, artistic creation, research, thoughtful analysis, problem posing, and problem solving? Let the tests ask them to

From *Educational Leadership*, vol. 46, no. 7, p. 41-47, 1989. Reprinted with permission.

write, speak, listen, create, do original research, analyze, pose and solve problems.

Rather than seeing tests as after-the-fact devices for checking up on what students have learned, we should see them as instructional: the central vehicle for clarifying and setting intellectual standards. The recital, debate, play, or game (and the criteria by which they are judged)—the "performance"—is not a checkup, it is the heart of the matter; all coaches *happily* teach to it. We should design academic tests to be similarly standard setting, not merely standardized.

Reform of testing depends, however, on teachers' recognizing that standardized testing evolved and proliferated because the school transcript became untrustworthy. An "A" in "English" means only that some adult thought the student's work was excellent. Compared to what or whom? As determined by what criteria? In reference to what specific subject matter? The high school diploma, by remaining tied to no standard other than credit accrual and seat time, provides no useful information about what students have studied or what they can actually do with what was studied.

To regain control over both testing and instruction, schools need to rethink their diploma requirements and grades. They need a clear set of appropriate and objective criteria, enabling both students *and outsiders* to know what counts, what is essential—what a school's standards really are. Until we specify what students must directly demonstrate to earn a diploma, they will continue to pass by meeting the de facto "standard" of being dutiful and persistent—irrespective of the quality of their work. And standardized testmakers will continue to succeed in hawking simplistic norm-referenced tests to districts and states resigned to using them for lack of a better accountability scheme.

Reform of testing depends on teachers' recognizing that standardized testing evolved and proliferated because the school transcript became untrustworthy.

EXHIBITIONS OF MASTERY

> The diploma should be awarded upon a successful final demonstration of mastery for graduation—an "Exhibition".... As the diploma is awarded when earned, the school's program proceeds with no strict age grading and with no system of "credits earned" by time spent in class. The emphasis is on the students' demonstration that they can do important things.
>
> —From the Prospectus of the Coalition of Essential Schools

The "exhibition of mastery," proposed by Ted Sizer in *Horace's Compromise* (1984) and a cornerstone of the "Essential School," is one attempt to grapple with these issues. The intent of the exhibitions project is to help schools and districts design more authentic, engaging, revealing, and trustworthy "tests" of a student's intellectual ability.

An "A" in "English" means only that some adult thought the student's work was excellent. Compared to what or whom? As determined by what criteria? In reference to what specific subject matter?

The reference to engagement is not incidental. The exhibition of mastery was initially proposed as an antidote to student passivity and boredom, not merely as a more valid form of assessment. The idea is to capture the interest value of an authentic test of one's ability, such as is often provided in schools by literary magazines, portfolios, recitals, games, or debates. Thus, "any exhibition of mastery should be the students' opportunity to show off what they know and are able to do rather than a trial by question..."[1]

The exhibition of mastery, as the name implies, is meant to be more than a better test. Like the thesis and oral exam in graduate school, it indicates whether a student has *earned* a diploma, is ready to leave high school.[2] The school is designed "backwards" around these standard-setting tests to ensure that teachers and students alike understand their obligations and how their own efforts fit in a larger context. Teachers "teach to the test" because the test is essential—and teacher designed.

The Rite of Passage Experience (R.O.P.E.) at Walden III , Racine, Wisconsin

All seniors must complete a portfolio, a study project on U.S. history, and 15 oral and written presentations before a R.O.P.E. committee composed of staff, students, and an outside adult. Nine of the presentations are based on the materials in the portfolio and the project; the remaining six are developed for presentation before the committee. All seniors must enroll in a yearlong course designed to help them meet these requirements.

The eight-part *portfolio*, developed in the first semester, is intended to be "a reflection and analysis of the senior's own life and times." The requirements include:

- a written autobiography,
- a reflection on work (including a resume),
- an essay on ethics,
- a written summary of coursework in science,
- an artistic product or a written report on art (including an essay on artistic standards used in judging artwork).

The *project* is a research paper on a topic of the student's choosing in American history. The student is orally questioned on the paper in the presentations before the committee during the second semester.

The *presentations* include oral tests on the previous work as well as six additional presentations on the essential subject areas and "personal proficiency" (life skills, setting and realizing personal goals, etc.). The presentations before the committee usually last an hour, with most students averaging about 6 separate appearances to complete all 15.

A diploma is awarded to those students passing 12 of the 15 presentations and meeting district requirements in math, government, reading and English.

Note: This summary is paraphrased from both the R.O.P.E. Student Handbook and an earlier draft of Archbald and Newmann's (1988) *Beyond Standardized Testing.*

Figure 1: An Example of a Final Exhibition

But why institute a radically new form of assessment? Why not just improve conventional teaching and course-related tests? As the "Study of High Schools" documented, a major cause of the high school's inadequacies is the absence of direct teaching of the essential skills of inquiry and expression. Even in "demanding" schools, students often fail to learn how to learn. The culprit is discipline-based curriculums that lead to content-based teaching and testing: the essential (cross-disciplinary) habits and skills of reading, writing, questioning, speaking, and listening fall through the cracks of typical content-focused syllabi

and course credits; as indicated, for example, when teachers say "I teach English, not reading."

A required final public exhibition of know-how ensures that those essentials are taught and learned. The final exit-level exhibition reveals whether a would-be graduate can demonstrate control over the skills of inquiry and expression and control over an intellectual topic that approximates the expert's ability to use knowledge effectively and imaginatively. A final exhibition provides students with an occasion to make clear, if only perhaps symbolically, that they are ready to graduate.

An exhibition challenges students to show off not merely their knowledge but their initiative; not merely their problem solving but their problem posing; not just their learning on cue, but their ability to judge and learn how to learn on an open-ended problem, often of their own design. The experience thus typically focuses on the essential skills of "inquiry and expression"—a synthesis that requires questioning, problem posing, problem solving, independent research, the creation of a product or performance, and a *public* demonstration of mastery. Significantly, there is often a component calling for self-reflection and analysis of what one has undergone and learned.

Thus, a *final exhibition* is a misnomer in an important sense. Many Coalition schools provide a semester- or yearlong course, an adult adviser, and a committee to ensure that a student has adequate guidance, evaluation, and incentive (see Fig. 1 for an example of a final exhibition from a Coalition school). The exhibition of mastery is as much a process as a final product, if not more so. The process of choosing topics, advisers, and committees and refining one's ideas and skills is a yearlong exercise in understanding and internalizing standards.

A similar approach to a diploma at the college level has been used successfully at Alverno College, Milwaukee, Wisconsin, for over a decade.[3] Assessment is a central experience, with coursework a means to a set of known ends: students must achieve mastery in the following eight general areas, with their progress in each area being charted on a multistaged scale:

1. effective communication ability,
2. analytic capability,
3. problem-solving ability,
4. valuing in a decision-making context,
5. effective social interaction,
6. taking responsibility for the global environment,
7. effective citizenship,
8. aesthetic responsiveness.

PERFORMANCES: BETTER CLASSROOM TESTS

Course-specific tests also have glaring weaknesses, not only because they are often too low level and content heavy. They are rarely designed to be authentic tests of intellectual ability; as with standardized tests, teacher-designed finals are usually intended to be quickly read and scored.

It seems wise, then, to talk about a move toward more intellectual performances in course-bound testing as a way of stressing the need to make tests more central, authentic, and engaging—as in the arts and athletics. (The term *exhibitions* would be reserved for those culminating graduation-level exercises designed to assess ability in the essentials underlying all coursework required for graduation.)

The high school diploma, by remaining tied to no standard other than credit accrual and seat time, provides no useful information about what students have studied or what they can actually do with what was studied.

Designing performances implies a very different approach to standard setting than is implied by typical criterion-referenced tests or outcome-based views of mastery, though the instincts behind the designs are similar. Performances would ideally *embody* and *evoke* desired outcomes in authentic contexts. Too often, specifying only outcomes leads to tests that atomize and decontextualize knowledge: the testmaker designs a set of isolated pat exercises designed to elicit each desired outcome. Genuine tests of ability rarely provide such blatant cues and simple recall; they require us to have a repertoire, the judgment and skill to "put it all together" in one central challenge, repeatedly tried. (Imagine the assessment of music ability in a series of little exercises tried once, rather than through practice and performance of a complete piece in recitals.)

An Oral History Project for 9th Graders

To the student:

You must complete an oral history based on interviews and written sources and then present your findings orally in class. The choice of subject matter is up to you. Some examples of possible topics include: your family, running a small business, substance abuse, a labor union, teenage parents, and recent immigrants.

Create three workable hypotheses based on your preliminary investigations and four questions you will ask to test out each hypothesis.

Criteria for Evaluation of Oral History Project

To the teacher:

Did student investigate three hypotheses?
Did student describe at least one change over time?
Did student demonstrate that he or she had done background research?
Were the four people selected for the interviews appropriate sources?
Did student prepare at least four questions in advance, related to the hypothesis?
Were those questions leading or biased?
Were follow-up questions asked where possible, based on answers?
Did student note important differences between "fact" and "opinion" in answers?
Did student use evidence to prove the ultimate best hypothesis?
Did student exhibit organization in writing and presentation to class?

Note: This example is courtesy of Albin Moser, Hope High School, Providence, Rhode Island. To obtain a thorough account of a performance-based history course, including the lessons used and pitfalls encountered, write to Dave Kobrin, Brown University, Education Department, Providence, RI 02912.

Figure 2: An Example of a Test Performance

In sum, the goals behind the exhibition of mastery and the performance are to design standard-setting tests that provide more direct evidence of a student's intellectual ability; design tests that are thus able to stand by themselves as objective results; design more authentic intellectual challenges at the heart of a discipline; and design tests that are more likely to engage students and motivate them to raise their own intellectual standards to do well on them. (See Fig. 2 for an example of a performance that illustrates and illuminates these design standards.)

TOWARD MORE AUTHENTIC TESTS

Exhibitions and performances sound fine on a schoolwide basis, you say, but districtwide or statewide? Isn't that too costly and cumbersome? I contend that the supposed impracticality and/or expense of designing such tests on a wide scale is a habit of thinking, not a fact. The United States is the only major country that relies so heavily on norm-referenced, short-answer tests instead of performance- and/or classroom-based assessment on a national level. In addition, a national committee on assessment in Great Britain has called for an exemplary system requiring flexible, criterion-referenced, and performance-based tests.[4] Many of the tests would be created by classroom teachers, who would be part of the standardizing process through "moderating" meetings to compare and balance results on their own and national tests.

In the U.S., more authentic skill assessment can now be found in various districts and states due, in part, to the work in writing assessment by the National Writing Project and its state offshoots (such as the California CAP writing test), and the American Council on the Teaching of Foreign Languages in the assessment of foreign language proficiency. Some states, such as Connecticut, have already designed and piloted performance-based assessment using ACTFL tests and criteria. In addition, they have piloted hands-on tests in graphics, small engines, and science. Vermont has proposed a statewide assessment system in writing and mathematics that would be portfolio based and teacher assessed.

We already have a national example in science: the 1987 NAEP pilot "Higher-Order Thinking Science Test," which includes some (though too few) hands-on experiments. One example:

> Students are given a sample of three different materials and an open box. The samples differ in size, shape, and weight. The students are asked to determine whether the box would weigh the most (and least) if it were completely filled with material A, B, or C. There are a variety of possible approaches...NAEP administrators used detailed checklists to record each student's procedures and strategies.[5]

A. Structure and Logistics

1. Are more appropriately public; involve an audience, a panel, and so on.
2. Do not rely on unrealistic and arbitrary time constraints.
3. Offer known, not secret, questions or tasks.
4. Are more like portfolios or a *season* of games (not one-shot).
5. Require some collaboration with others.
6. Recur—and are *worth* practicing for, rehearsing, and retaking.
7. Make assessment and feedback to students so central that school schedules, structures, and policies are modified to support them.

B. Intellectual Design Features

1. Are "essential"—not needlessly intrusive, arbitrary, or contrived to "shake out" a grade.
2. Are "enabling"—constructed to point the student toward more sophisticated use of the skills or knowledge.
3. Are contextualized, complex intellectual challenges, not "atomized" tasks, corresponding to isolated "outcomes."
4. Involve the student's own research or use of knowledge, for which "content" is a means.
5. Assess student habits and repertoires, not mere recall or plug-in skills.
6. Are *representative* challenges—designed to emphasize *depth* more than breadth.
7. Are engaging and educational.
8. Involve somewhat ambiguous ("ill-structured") tasks or problems.

C. Grading and Scoring Standards

1. Involve criteria that assess essentials, not easily counted (but relatively unimportant) errors.
2. Are not graded on a "curve" but in reference to performance standards (criterion-referenced, not norm-referenced).
3. Involve demystified criteria of success that appear to *students* as inherent in successful activity.
4. Make self-assessment a part of the assessment.
5. Use a multifaceted scoring system instead of one aggregate grade.
6. Exhibit harmony with shared schoolwide aims—a *standard*.

D. Fairness and Equity

1. Ferret out and identify (perhaps hidden) strengths.
2. Strike a *constantly* examined balance between honoring achievement and native skill or fortunate prior training.
3. Minimize needless, unfair, and demoralizing comparisons.
4. Allow appropriate room for student learning styles, aptitudes, and interests.
5. Can be—should be—attempted by *all* students, with the test "scaffolded up," not "dumbed down," as necessary.
6. Reverse typical test-design procedures: they make "accountability" serve student learning (Attention is primarily paid to "face" and "ecological" validity of tests).[1]

1. Thanks to Ted Sizer, Art Powell, Fred Newmann, and Doug Archbald; and the work of Peter Elbow and Robert Glaser for some of these criteria.

Figure 3: Characteristics of Authentic Tests

NAEP borrowed most of its experiments from the British Assessment of Performance Unit tasks, which have been used (and reliably scored) in Great Britain for a decade in reading, speaking, listening, math, and science.

Genuine tests *can* be widely implemented if we can overcome inertia and fatalism about current forms of standardized testing. Authentic, performance-based testing is a reality, not a romantic vision. There is also ample room for more intelligent design and use of conventional norm-referenced standardized tests.[6]

The state of Connecticut has developed a "Common Core of Learning," which lists objectives and criteria in all essential domains. Performance-based tests, built around criteria specified by experts in each field and involving tests administered by trained observers, are to be designed to honor those aims.

There are even standardized tests worth noting. ACT has developed a wide-ranging multimedia test of "general education knowledge and skills" called COMP, designed for colleges but easily adaptable to the high school level. The test uses art reproductions and audiotapes of news programs, for example, in testing writing and listening skills. On other items, students draft letters on different topics. There is even allowance for the student to respond orally on tape to a few test questions. The test takes six hours to administer, covers all the essential skills of inquiry and expression, and includes a 54-question self-assessment about one's patterns of activity related to each competency.

To talk with disdain of "teaching to the test" is to misunderstand how we learn.

In sum, authentic tests have four basic characteristics in common. First, they are designed to be truly representative of performance in the field; only then are the problems of scoring reliability and logistics of testing considered. Second, far greater attention is paid to the teaching and learning of the *criteria* to be used in the assessment. Third, self-assessment plays a much greater role than in conventional testing.[7] And, fourth, the students are often expected to present their work and defend themselves publicly and orally to ensure that their

apparent mastery is genuine. (See Fig. 3 for a more thorough list of characteristics of authentic tests.)

TOWARD A PERFORMANCE-BASED DIPLOMA

The diploma by exhibition implies radically different standards for graduation. Instead of seat time or the mere accrual of Carnegie units, the diploma is performance based and criterion referenced. We may not be ready for the demise of age grading and social promotion; but if the harm done by standardized testing is to be undone, we need to redesign schools "backwards" around graduation-level standards of performance.

The performances and exhibitions should be designed prior to instruction, thus setting the school's standards in functional, not merely abstract and idealized, terms. Seeing them as add-ons to the traditional curriculum is to miss the point. How must the school be redesigned to support exhibitions or any form of exit-level standards? This should be the question behind "restructuring" and the source of vigorous debate among faculties and school board members. Designing and institutionalizing exhibitions would better ensure, in other words, that the school had clear, coherent, and effective standards. Knowing the desired student abilities and work standards, as embodied in culminating performances and scoring criteria, would force key issues of policy: how will time, space, personnel, and other resources be best spent to ensure that diploma standards are met?

To talk with disdain of "teaching to the test" is to misunderstand how we learn. The test is the point of leverage—for learning and for reform. The issue is the integrity of the test: the genuineness, effectiveness, and aptness of the challenge. The finals (and the criteria by which they are graded) set the standards of acceptable work in a course and a school—irrespective of noble language in school district reports or teacher intentions as reflected in syllabi. Legitimate and effective assessment is as simple(!) as ensuring that tests, grades, diploma requirements, and the structures and policies of the schools practice what we preach as essential. If we so honor our professed aims, the problems associated with standardized testing will take care of themselves.

NOTES

1. From *Horace's Compromise* (Sizer 1984), p. 68.
2. This (final) exhibition is patterned after the 18th century model of a public display of one's ability to engage in disputation: "... candidates for degrees expected to be academically tested at Commencement itself. Bachelor of Arts candidates prepared theses or topics on which they could be quizzed, and candidates for Master of Arts submitted questions they were ready to defend. Titles of theses and questions were printed in advance and handed out at Commencement, and visitors often took the opportunity of challenging the candidates on their knowledge" (from the Harvard University Commencement program).
3. See the booklet *Assessment at Alverno College*, available from the college. For a history and an analysis of Alverno's program (as well as a general discussion of competency-based higher education), see *On Competency* (Grant, Elbow et al. 1979).
4. *National Curriculum: Task Group on Assessment and Testing: A Report.* Available from the Department of Education and Science (1988). This is a landmark document, outlining in readable prose a plan for intelligent and humane assessment.
5. From *Learning By Doing* (Educational Testing Service 1987).
6. See the excellent article by Dan Koretz of the RAND Corporation in the Summer 1988 issue of *American Educator*, which sums up the current controversy about norm-referenced state testing (the "Lake Wobegon effect" of each state being above average) and provides a useful set of guidelines for assessing assessment.
7. At Alverno, self-assessment is often the first level of proficiency. Thus, in the speaking requirement, students must give a five-minute videotaped talk—with the first evaluations given on the student's self-assessment after watching the videotape.

REFERENCES

Alverno College Faculty. (1979/1985). *Assessment at Alverno College.* Rev. ed. Milwaukee, Wis.: Alverno College.

Archbald, D., and F. Newmann. (1988). *Beyond Standardized Testing: Authentic Academic Achievement in the Secondary School.* Reston, Va.: NASSP Publications.

Department of Education and Science and the Welsh Office. (1988). *National Curriculum: Task Group on Assessment and Testing: A Report.* London: Her Majesty's Stationery Office, Department of Education and Science, England and Wales. A brief "Digest for Schools" is also available.

Educational Testing Service. (1987). *Learning by Doing: A Manual for Teaching and Assessing Higher-Order Thinking in Science and Mathematics.* A report on the NAEP pilot of performance-based assessment. A summary of the NAEP pilot of performance-based assessment. Princeton, N.J.: ETS. The full report: *A Pilot Study of Higher-Order Thinking Skills Assessment Techniques in Science and Mathematics.* ETS Report #17-HOS-80.

Grant, G., P. Elbow, et al. (1979). *On Competence: A Critical Analysis of Competence-Based Reforms in Higher Education.* San Francisco: Jossey-Bass.

Koretz, D. (Summer 1988). "Arriving in Lake Wobegon: Are Standardized Tests Exaggerating Achievement and Distorting Instruction?" *American Educator 12*, 2.

Sizer, T. (1984). *Horace's Compromise: The Dilemma of the American High School.* Updated ed. Boston: Houghton-Mifflin.

Wiggins, G. (In press). "A True Test: Toward Authentic and Equitable Forms of Assessment." *Phi Delta Kappan.*

RECOMMENDED READINGS

Alverno College Faculty. (1984). *Analysis and Communication at Alverno: An Approach to Critical Thinking.* Milwaukee, Wis.: Alverno College.

Berk, R. A., ed. (1986). *Performance Assessment: Methods and Applications.* Baltimore, Md.: Johns Hopkins University Press.

Bloom, B., G. Madaus, and J. T. Hastings. (1981). *Evaluation to Improve Learning.* New York: McGraw-Hill.

Brooks, G. (1987). *Speaking and Listening: Assessment at Age 15.* Great Britain: The Assessment of Performance Unit (APU), Department of Education and Science. APU material exists on the results of performance-based assessment in language, history, math, science, and history in primary and secondary schools.

Elbow, P. (1986). "Trying to Teach While Thinking About the End and Evaluating Students More Accurately." In *Embracing Contraries: Explorations in Teaching and Learning.* New York: Oxford University Press. The former chapter originally published in Grant, Elbow, et al. (1979).

Higgs, T., ed. (1984). *Teaching for Proficiency, The Organizing Principle.* Lincolnwood, Ill.: National Textbook Company and ACTFL.

Sizer, T. (1986). "Changing Schools and Testing: An Uneasy Proposal." In *The Redesign of Testing for the 21st Century.* 1985 ETS Invitational Conference Proceedings. Princeton, N.J.: ETS.

Slavin, R., et al. (1986). *Using Student Team Learning.* 3rd ed. Baltimore: The Johns Hopkins Team Learning Project Press.

Snow, R. (1988). "Progress in Measurement, Cognitive Science, and Technology That Can Change the Relation Between Instruction and Assessment. In *Assessment in the Service of Learning.* 1987 ETS Invitational Conference Proceedings. Princeton, N.J.: ETS.

Spandel, V. (1981). *Classroom Applications of Writing Assessment: A Teacher's Handbook.* Portland, Oreg.: Northwest Regional Educational Laboratory.

Stiggins, R. (1987). "Design and Development of Performance Assessments." *Educational Measurement: Issues and Practices*, 6, 3: 33-42. An Instructional Model (ITEMS), published by National Council on Measurement in Education (NCME). Comes with an Instructor's Guide.

Stiggins, R. (January 1988). "Revitalizing Classroom Assessment." *Phi Delta Kappan*, 69: 5.

Wiggins, G. (Winter 1987). "Creating a Thought-Provoking Curriculum." *American Educator*, 11, 4.

Wiggins, G. (Winter 1988). "Rational Numbers: Scoring and Grading That Helps Rather Than Hurts Learning." *American Educator*, 12, 4.

Alternatives to Testing

Three categories of assessment methods are available to gauge student attainment of our expectations: paper-and-pencil *assessments,* performance *assessments (or measures based on observation and judgment), and* direct personal communication *with the students.*–Richard J. Stiggins

If we are going to reduce the volume of standardized tests we currently administer and de-emphasize test results, then how can we measure accurately whether or not our students are learning? Attention is shifting from the tremendous emphasis we place on norm-referenced tests to alternative forms of assessments that provide more opportunities for students to demonstrate what they can do rather than recall what they know. Steven Ferrara and Jay McTighe's photographic analogy shows how we often take a candid photograph of a person and no matter how unflattering or inappropriate, that photo or one-time snapshot represents that person. Often, that one-time photo for a student is the standardized test. Despite the bad day he might have had or the outside variables (testing conditions, test phobia, test bias, personal problems) that might have beset him, that test score is often the determining factor to measure what the student has learned. That one photo could also determine his placement in future classes (tracking or streaming), his acceptance into college, or his future employment opportunities.

Is it fair to judge a student by a single photograph when teachers could, in the words of Ferrara and McTighe, "construct a 'photo album' containing a variety

of pictures taken at different times with different lenses, backgrounds, and compositions"?

Effective classroom assessment requires a wide variety of tools gathered over a long period of time. Effective assessment can help adjust for the learning styles of students, multicultural differences, and learning and behavior disabilities. Effective assessment also enhances the curriculum by allowing students some *choice* in what they want to do and how they want to do it. Carolyn Mamchur advocates giving students choices in deciding what topics they want to study for special projects, choosing the order they want to study things, and "matching students' learning styles, interests, and abilities to a task."

Instead of the traditional tools of standardized test scores, teacher-made test scores, quizzes, worksheets, papers, and homework, teachers can provide students with a wide range of options for assessment. By using a repertoire of assessment tools, teachers can allow students to express themselves according to their individual learning style and, according to Fred M. Newmann, "to engage in discipline inquiry to produce knowledge that has value in their lives beyond simply proving their competence in school."

Some of the authentic assessment tools suggested in this collection of articles include the following:

- logs
- journals
- writing samples
- performances
- observations
- interviews
- conferences
- projects
- videos and cassettes
- computer programs
- kid watching
- questioning
- oral presentations
- labs
- exhibitions
- portfolios
- self-evaluations

The data derived from this repertoire of tools provides a clearer picture of students' strengths and weaknesses. In addition, this method allows students to *choose* how to demonstrate their mastery of significant learning outcomes.

If we are to become "assessment literates" as Stiggins defines the term, we need to seek and use assessments that communicate clear and rich pictures of the authentic achievement and the significant outcomes that students need as lifelong learners.

Our task, according to Newmann, is to help students produce, rather than reproduce knowledge, and our "'restructured' vision of the goals of education seeks to evaluate performance activities that are worthwhile, significant, and meaningful: in short, activities that are authentic."

Searching for Alternatives to Standardized Tests: Whys, Whats, and Whithers

by Walter Haney and George Madaus

More than new forms of assessment, what is needed is a refusal to accept bondage to any single technology and an assertion of the professional right to choose different kinds and mixes of assessments for different purposes, Messrs. Haney and Madaus assert.

THE VOLUME of standardized testing in the nation's schools has been increasing steadily over the last several decades. At the same time, long-standing complaints about standardized tests have grown louder. This growing dissatisfaction has led to a search for alternative ways to measure the learning of students, the performance of the nation's schools, and the employment potential of adults.

In this article we review developments in this area and try to answer these questions: How much has standardized testing been increasing? What is the case against such testing, and why are people searching for alternatives? What alternatives are available? Whither might they lead us?

We believe that the search for alternatives is somewhat shortsighted. Whether the coin of assessment and evaluation in the schools is standardized tests or some alternative, what is needed—more than alternative instruments—is a critical attitude toward assessment and a wider appreciation of its effects on teaching and learning.

From *Phi Delta Kappan*, vol. 70, no. 9, p. 683-687, May 1989.
Reprinted with permission.

THE UPSURGE IN TESTING

Standardized tests have been used in U.S. schools for more than a hundred years. Before the turn of the century, however, most of the tests were what would now be called criterion-referenced tests: tests of skills in such subjects as spelling, writing, and arithmetic that are scored with reference to examples of poor, fair, and good performance in those skills. After the development of norm-referenced tests—which compare any given student's performance with that of other test-takers—the use of nationally normed achievement testing became increasingly common in the schools. The popularization of intelligence testing, as a result of the much-publicized use of the Army Alpha and Army Beta tests during World War I, contributed to a dramatic expansion of norm-referenced testing in the nation's schools in the 1920s and 1930s. By the 1930s the Stanford Achievement Test alone was selling at a rate of about 1 1/2 million booklets per year.

The next large jump in the use of testing in the schools took place during the 1950s and 1960s, largely as a result of federal legislation that promoted testing (notably the National Defense Education Act of 1958 and the Elementary and Secondary Education Act of 1965). In the 1980s testing seems to have received a substantial boost from a variety of initiatives focused on the reform of education. According to Chris Pipho of the Education Commission of the States, "Nearly every large education reform effort of the last few years has either mandated a new form of testing or expanded uses of existing testing."[1]

No precise estimates of the volume of testing in the schools are available, in large measure because so many different tests are used for so many different purposes by so many different sponsoring agencies. However, by piecing together indirect evidence from a variety of sources, we have estimated that the volume of testing in the schools has been increasing by between 10% and 20% annually over the last 40 years (with the higher estimate relevant for more recent years).[2] To cite just one piece of indirect evidence, the average annual number of column inches devoted to articles about the curriculum in the *Education Index* has increased only modestly over the last

50 years (see graph at right)—from between 50 and 100 column inches in the 1930s and 1940s to between 100 and 150 column inches in recent years. In sharp contrast, the average annual number of column inches devoted to articles about testing in the *Education Index* has increased dramatically—from between 10 and 30 inches in the 1930s and 1940s to well over 300 inches in the 1980s. This is only one of many pieces of evidence that suggest to us that standardized testing has become the coin of the realm in education.

COMPLAINTS

However, as the quantity of testing in the nation's schools has risen, complaints about standardized tests have increased in both volume and intensity.[4] In addition to published complaints, there have been hearings in which standardized testing has been subjected to criticism—in Congress and in several state legislatures—and a "truth-in-testing" law has been passed in New York. In addition, since the 1960s there have been hundreds (if not thousands) of lawsuits brought in both state and federal courts concerning testing in education and in employment. Indeed, the complaints, criticisms, and caterwauling have been so substantial that some have written of a "war on testing."[5]

As the quantity of testing in the nation's schools has risen, complaints about standardized tests have increased in both volume and intensity.

We do not intend to provide a play-by-play commentary on the recent controversies over standardized testing, nor do we intend to name all the major players and their various plaints and ploys. Instead, we simply classify several decades' worth of complaints against standardized testing into four categories of recurring criticisms. The critics charge that standardized tests:

- give false information about the status of learning in the nation's schools;
- are unfair to (or biased against) some kinds of students (e.g., minority students, those with limited proficiency in English, females, and students from low-income families);

- tend to corrupt the processes of teaching and learning, often reducing teaching to mere preparation for testing; and
- focus time, energy, and attention on the simpler skills that are easily tested and away from higher-order thinking skills and creative endeavors—the Achilles' heel of the nation's education system today, in the view of many observers.

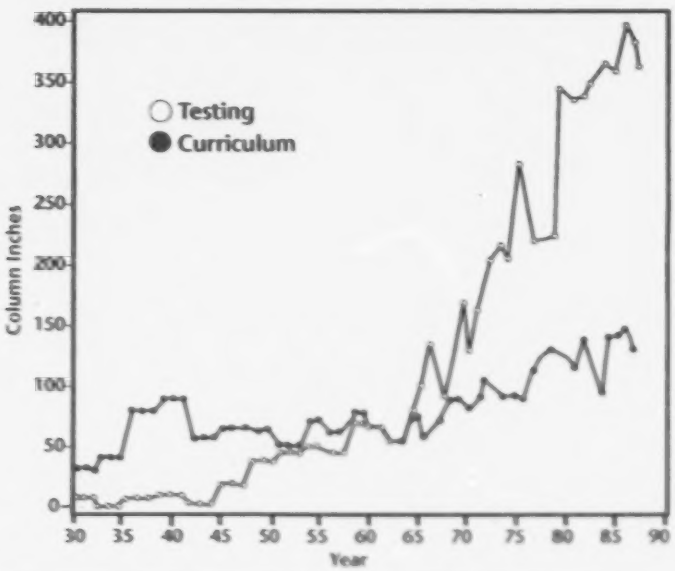
THE SEARCH FOR ALTERNATIVES

Given the seriousness and range of such charges, it is not surprising that many people have been searching for alternatives to standardized tests in educational and employment settings. In the 1970s, for example, when the National Education Association (NEA) and other education organizations began to criticize publicly the influence of standardized tests on the schools, there was widespread interest in alternatives. The NEA passed a resolution encouraging the "elimination of group standardized intelligence, aptitude, and achievement tests" and included it in the handbook titled *Alternatives to Standardized Testing*.[6] Among the suggested alternatives to standardized, norm-referenced tests were the professional judgment of teachers, samples of student work, contracts with students, interviews, teacher-made tests, and criterion-referenced tests.

In employment, interest in alternatives was prominently promoted by a set of stiff guidelines issued in 1970 by the U.S. Equal Employment Opportunity Commission (EEOC) to prevent the discriminatory use of employment tests. The guidelines specified that the use of any test that adversely affected the hiring, promotion, or other employment opportunities of classes protected under Title VII of the Civil Rights Act constituted discrimination in employment unless the test had been validated to show its relevance to the job and unless it had been demonstrated that alternative selection procedures with less discriminatory impact were unavailable.

The EEOC guidelines were revised in 1978, and their interpretation has changed considerably as a result of changing Administrations and evolving judicial interpreta-

***Education Index* Listings Under Testing and Curriculum**



tions. However, it is clear that the guidelines have focused considerable attention on alternatives to cognitive tests as selection instruments in employment. Richard Reilly and Georgia Chao, for example, reviewed eight alternatives to paper-and-pencil tests as employment selection instruments, including biographical data, interviews, peer evaluations, self-assessments, reference checks, grades, and expert judgment.[7]

In addition to this longstanding interest in alternatives in both education and employment, two more general developments—the emerging field of cognitive science and progress in computer and information technology—have heightened interest in the possibility of finding alternatives to traditional standardized tests. A national conference on testing, held in 1979, concluded that advancements in cognitive science—an interdisciplinary movement involving psychologists, computer scientists, philosophers, linguists, anthropologists, and educators—could lead to new and better tests through a "powerful new understanding of the processes in learning

and performing the intellectual tasks to be tested."[8] Similarly, in 1982 the Committee on Ability Testing of the National Research Council called for greater use of cognitive science in test development:

> Theories of cognition...do not currently play a central role in test development. As a result the task of explaining even so long-recognized an ability as reading comprehension in terms of a theory of information processing or other advanced psychological concepts has barely begun. If more psychologists who worked on theories of cognition also worked on test development, testing today might be further advanced. This kind of research is difficult, but the effort is needed.[9]

The other broad development that has heightened hopes for new and better alternatives to standardized tests is the astounding pace of progress in computer and information technology. Computerized adaptive testing, in which performance on the first dozen or so items determines whether subsequent items are either easier or more difficult, is already being tried out. The theoretical advantages of computerized adaptive testing are such that "some advocates believe that within 15 years this form of testing will completely displace paper-and-pencil tests in many programs."[10]

Standardized tests became popular in the early part of this century, when school populations were growing rapidly and large school systems were being organized.

More broadly, many people believe that information technology now "provides a means to improve instruction and testing in one of the ultimate aims of education—the ability to address more complex problems of the sort encountered in work and in personal life."[11] This vision suggests that testing might take place at a video terminal that would present the student with a simulation requiring open-ended decision making of the sort that might be encountered in real life. In other words, computer and other information technology might break the stranglehold that the multiple-choice format has had on large-scale testing programs for 50 years or more.[12]

WHITHER THE ALTERNATIVES?

A wide range of alternatives to standardized tests has been suggested in the past, and even more alternatives have been predicted for the future. The continuing interest in alternatives suggests that the make-a-better-mousetrap mentality is alive and well in assessment circles in education and employment. We suggest exercising caution about the potential impact of most such alternatives, until there has been a careful consideration of the connections between any educational assessment and teaching and learning in the schools.

A point worth noting about evaluation alternatives that have been suggested over the last 20 years is that many of them are not at all new. Evaluation tools such as live performances, products, teacher judgment, and school grades have a long history in education. Standardized tests became popular in part because of problems with these older forms of evaluation. In the 19th century, for example, written exams were instituted in high schools in Boston because visiting committees did not have enough time to hear the recitations and performances of the increasing number of students.

In this century, the spread of standardized tests in the schools was aided by research showing that grading by teachers (e.g., of student essays) is highly unreliable. Moreover, standardized tests were suggested as alternatives to prevailing grading practices some 30 years ago, amid widespread concern about the ill effects of letter grades on high school students.[13] James Crouse and Dale Trusheim have shown that Scholastic Aptitude Test Scores contribute little, if any, predictive power to high school grades in helping colleges and universities make admissions decisions.[14] But college admissions tests are still used widely, at least in part (it seems to us) because they constitute uniform alternatives to reliance on high school records.

It is also worth noting that the customs and social organization of schools will certainly affect the inroads that alternatives can make on assessment. Standardized tests became popular in the early part of this century, when school populations were growing rapidly and large

school systems were being organized in increasingly bureaucratic fashion. Indeed, the term *educational testing* is something of a malapropism, since most standardized testing has far less to do with the teaching of individual teachers and the learning of individual students than it does with the bureaucratic organization of schools.[15] Thus new testing technologies that seem to have real potential for helping to integrate testing and teaching may be hurt in the marketplace because they lack the potential of traditional norm-referenced tests to enforce bureaucratic accountability.

Promising new testing technologies on the horizon could potentially broaden the range of human talent that is assessed. But the prospect of technological solutions to educational problems remains highly doubtful.

Consider the case of the Degrees of Reading Power (DRP) test. A decade ago the DRP was cited in a federal report on testing as an example of work in psycholinguistics that held much promise for helping to integrate testing with teaching.[16] Moreover, independent reviews suggest that the DRP test does indeed have some real strengths in comparison with traditional norm-referenced tests of reading.[17] Yet the DRP appears to have had only limited success in the marketplace. For example, it is being dropped this year from the systemwide testing program in Boston, and as of June 1989 the College Board will no longer market the instrument (although it will still be available from the developer). One of the early complaints that many school administrators voiced about the DRP was that it lacked the national norms that would show how their schools were performing in comparison to schools nationwide. The norms now have been developed.

Another general caveat about alternatives to standardized tests arises from what so far appears to have been the impact of computer technology on large-scale testing programs. At present, the major effect of computer technology on testing has been in the creation of more efficient multiple-choice tests rather than in the exploitation of computer technology to create real alternatives to standardized multiple-choice tests.[18]

In sum, it seems to us that some promising new testing technologies on the horizon could potentially broaden the range of human talent that is assessed. But the prospect of technological solutions to educational problems remains highly doubtful. Note, for example, that Howard Gardner and Robert Sternberg have received considerable attention for their recent work on alternative forms of intelligence, including nonacademic intelligence. Their work has frequently been cited as a likely source of alternatives to standardized tests.[19] Yet the prevailing form of assessment changed little when such people as John Holland and Leonard Baird more than 20 years ago showed the potential importance of and provided the means for assessing nonacademic achievements. Thus, in addition to exploring new technologies of testing and assessment, we recommend paying much closer attention to the effects on education of whatever forms of assessment are employed and maintaining a healthy skepticism toward any single technology of assessment.

On the first point, we cite an investigation into the effects that assessment through objective tests and through writing samples have had on instruction in composition. Charles Suhor sought both research data and personal opinion, the latter through informal polling of 350 state and local supervisors of English and language arts. From his informal polling, Suhor concluded: "With few exceptions the testimony on objective tests was consistent and damning: both standardized, norm-referenced and criterion-referenced competency tests discourage writing instruction. . . . By contrast, praise was abundant for writing samples. . . ."[20]

The influence of testing is to reinforce the objectives sought by the school. The *form* of assessment should also faithfully reflect educational intentions.

Suhor could find little on the topic in the research literature, but what he found "essentially supports the overwhelming testimony of practitioners."[21] Grading of writing samples is surely more time-consuming and less efficient than grading multiple-choice tests of writing skills, but it has the tangible benefit of directing attention toward actual writing. As Ralph Tyler wrote 30 years ago,

"Since tests have a powerful directive influence on teaching and the study of pupils, a major policy to follow is to establish a testing program that faithfully reflects the objectives sought by the school. In this way the influence of testing is to reinforce the objectives sought by the school."[22] We would add that the *form* of assessment should also faithfully reflect educational intentions.

The second point we wish to make about alternative assessments is that *what* technology of assessment is used probably makes far less difference than *how* it is used. As with other forms of technology touted for schools in the past, technologies of assessment probably have far less potential for changing schools—for good or ill—than many of the controversies surrounding testing suggest. As Stanley Reiser advised in his history of technological innovation in medicine, *how* a tool is used is often more important than *which* tool is used:

> If physicians in general come to accept a fundamentally mechanical view of human beings, in a world that is more and more enamored of technology, the prospect for the future is extremely disquieting.... Machines inexorably direct the attention of both doctor and patient to the measurable aspects of illness, but away from the "human factors" that are at least equally important.... Technologies that improve accuracy, and centralized organizations that enhance efficiency and improve security, are essential factors in modern medicine. Yet accuracy, efficiency, and security are purchased at a high price when that price is impersonal medical care and undermining the physician's belief in his own medical powers. To be free to develop his medical skills to their highest point, to increase what is despite these problems a positive balance of benefits over harms, today's physician must rebel. He can use his strongest weapon—a refusal to accept bondage to any one technique, no matter how useful it may be in a particular instance. He must regard them all with detachment as mere tools, to be chosen as necessary for a particular task. He must accept the patient as a human being and regain and reassert high faith in his own medical judgment.[23]

If there is a general case against standardized multiple-choice tests, it is this: they are becoming so pervasive in the schools and the same tests are being used for so

many different purposes—from diagnoses of student learning to program evaluation, from teacher evaluation to systemwide accountability—that both test results and curriculum and instruction are becoming warped, if not corrupted.[24]

Alternatives of the sort widely proposed may help provide conditions that will allow teachers, parents, and others to regain and reassert faith in their own educational judgment, but technologies of educational assessment (of whatever variety) in and of themselves will never cure the ills commonly attributed to standardized tests. More than new forms of assessment, what is needed is a refusal to accept bondage to any single technology—no matter how useful it may be in a particular instance—and thoughtful selection of different kinds and mixes of assessments for different purposes.

NOTES

1. Chris Pipho, "Tracking the Reforms, Part 5: Testing—Can It Measure the Success of the Reform Movement?," *Education Week*, 22, May 1985, p. 19.
2. George Madaus and Walter Haney, *The Fractured Marketplace for Standardized Testing* (Berkeley, Calif.: National Commission on Testing and Public Policy, forthcoming).
3. Walter Haney, "Validity, Vaudeville, and Values: A Short History of Social Concerns Over Standardized Testing," *American Psychologist*, pp. 1021-34; idem, "Testing Reasoning and Reasoning About Testing," *Review of Educational Research*, Winter 1984, pp. 597-654; and Madaus and Haney, op. cit.
4. Banesh Hoffman, *The Tyranny of Testing* (New York: Crowell-Collier, 1962); Martin Gross, *The Brain Watchers* (New York: Random House, 1962); Vance Packard, *The Naked Society* (New York: McKay, 1964); Robert L. Williams, Doris Mosby, and Vincent Hinson, "Critical Issues in Achievement Testing of Children From Diverse Educational Backgrounds," paper presented at the Invitational Conference on Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation, U.S. Office of Education, Washington, D.C., 1976; Paul Houts, ed., *The Myth of Measurability* (New York: Hart Publishing, 1971); Allan Nairn, and Ralph Nader, *The Reign of ETS: The Corporation That Makes Up Minds* (Washington, D.C.: Public Citizen, 1980); Stephen J. Gould, *The Mismeasure of Man* New York: Norton, 1981); Andrew J. Strenio, Jr., *The Testing Trap* (New York: Rawson, Wade, 1981); David Owen, *None of the Above: Behind the Myth of the SAT* (Boston: Houghton Mifflin, 1895); James Crouse and Dale Trusheim, *The Case Against the SAT* (Chicago: University of Chicago Press, 1988); George Madaus, "The Influence of Testing on the Curriculum," in Laurel Tanner, ed., *Critical Issues in Curriculum: 87th NSSE Yearbook* (Chicago: National Society for the Study of Education, University of Chicago Press, 1988),pp. 83-121; John Cannell, *Nationally*

BEST COPY AVAILABLE

Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above the National Average (Daniels, W. Va.: Friends for Education, 1987); and idem, "The Lake Wobegon Effect Revisited," *Educational Measurement: Issues and Practice*, Winter 1988, pp. 12-15.

5. Barbara Lerner, *The War on Testing: Detroit Edison in Perspective* (Princeton, N.J.: Educational Testing Service, 1979); idem, "The War on Testing: David, Goliath, and Gallup." *Public Interest*, vol. 60, 1980, pp. 119-47; and Fred Hargadon, *Two Cheers* (Princeton, N.J.: College Board Publications Office, 1980).
6. Frances Quinto, and Bernard McKenna, *Alternatives to Standardized Testing* (Washington, D.C.: National Education Association, 1977), p. 7.
7. Richard Reilly and Georgia Chao, "The Validity and Fairness of Some Alternative Employee Selection Procedures," *Personnel Psychology*, vol. 35, 1982, pp. 1-62.
8. Ralph Tyler and Sheldon White, *Testing, Teaching, and Learning: Report of a Conference on Research on Testing* (Washington, D.C.: National Institute of Education, 1979), p. 5.
9. Alexandra Wigdor and Wendell Garner, eds., *Ability Testing: Uses, Consequences, and Controversies, Vol. I* (Washington, D.C.: National Academy Press, 1982), p. 214.
10. James R. McBride, "Computerized Adaptive Testing," *Educational Leadership*, October 1985, p. 25.
11. Tyler and White, p. 16.
12. George Madaus, "The Perils and Promises of New Tests and New Technologies: Dick and Jane and the Great Analytical Engine?," in *The Redesign of Testing for the 21st Century: Proceedings of the 1985 ETS Invitational Conference*. (Princeton, N.J.: Educational Testing Service, 1986), pp. 87-101.
13. Howard Kirschenbaum, Sidney Simon, and Rodney Napier, *Wad-Ja-Get?: The Grading Game in American Education*. (New York: Hart, 1971).
14. Crouse and Trusheim, op. cit.
15. Walter Haney, "Making Testing More Educational," *Educational Leadership*, October 1985, pp. 4-13.
16. Tyler and White, op. cit.
17. Robert Linn, "Issues of Validity for Criterion-Referenced Measures," *Applied Psychological Measurement*, Fall 1980, pp. 547-61; Michael Kibby, "Test Review: The Degrees of Reading Power," *Journal of Reading*, February 1981, pp. 416-27; and Haney, "Making Testing."
18. McBride, p. 26; and Madaus, "The Perils and Promises...."
19. Robert Sternberg, *Beyond I.Q.* (New York: Cambridge University Press, 1985); idem, *CAT: A Program of Comprehensive Abilities Testing*. (Berkeley, Calif.: National Commission on Testing and Public Policy, forthcoming); Howard Gardner, *Art, Mind, and Brain*. (New York: Basic Books, 1982); idem, *Frames of Mind: The Theory of Multiple Intelligences*. (New York: Basic Books, 1983); and idem, *Assessment in Context: The Alternative to Standardized Testing* (Berkeley, Calif.: National Commission on Testing and Public Policy, forthcoming).
20. Charles Suhor, "Objective Tests and Writing Samples: How Do They Affect Instruction in Composition?," *Phi Delta Kappan*, May 1985, p. 636.
21. Ibid., p. 639.

22. Ralph Tyler, "What Testing Does to Teachers and Students," in Anne Anastasi, ed., *Testing Problems in Perspective* (Washington, D.C.: American Council on Education, 1966) p. 49.
23. Stanley J. Reiser, *Medicine and the Reign of Technology.* (New York: Cambridge University Press, 1979), pp. 229-31.
24. Linda Darling-Hammond and Arthur Wise, "Beyond Standardization: State Standards and School Improvement," *Elementary School Journal*, January 1985, pp. 315-36.

Assessment Literacy

by Richard J. Stiggins

We allocate virtually no resources to train practitioners or interested others in sound methods of assessing learning outcomes, Mr. Stiggins avers. It's time, he believes, to change all that.

OVER THE PAST decade, educators and the public alike have become keenly aware of the importance of educational outcomes and high achievement standards. No longer is it sufficient, for purposes of "accreditation," that educators simply build the proper facilities, buy the right textbooks, maintain the proper student/teacher ratios, and have enough books in the library. Today, such *process* variables are to be used in the right combinations to produce the desired *product*: measured student learning.

Evidence abounds of our newfound devotion to outcomes. Virtually all local education agencies conduct districtwide testing programs. Nearly every state conducts a statewide testing program. Local and national news media devote major coverage to the presentation of outcome data, such as changing patterns of test scores over time. The U.S. Department of Education annually reports state-by-state measures of a variety of outcomes. International comparisons of academic achievement are becoming more frequent and more visible.

No longer is it sufficient, for purposes of "accreditation," that educators simply build the proper facilities, buy the right textbooks, maintain the proper student/teacher ratios, and have enough books in the library.

Important instructional and policy decisions are based on information about student achievement. If those

From *Phi Delta Kappan*, vol. 72, no. 7, p. 534-539, March 1991. Reprinted with permission.

decisions are made well, students benefit. If they are made poorly, students suffer. But those decisions can only be as sound as the data on which they are based.

Are reliable data available? The answer depends on the particular political, educational, and assessment context within which we seek the data. In some contexts, high-quality outcome data abound; in others, not so. Decision makers must know when they have sound data in hand and when they do not. This is where we come face to face with a critical problem: most decision makers—educators and noneducators alike—are not sufficiently literate in the basics of assessment to know whether their achievement data are sound or unsound. We are a nation of assessment illiterates. We are a society that has come to care very much about high standards of achievement, but we are a society that is incapable of understanding whether those standards are being met.

WHAT IS ASSESSMENT LITERACY?

Assessment literacy is best understood by looking first at what it means *not* to have it. Assessment illiterates don't understand what it takes to produce high-quality achievement data and so do not evaluate critically the data they use. Assessment illiterates accept achievement data at face value and can easily be intimidated by apparently technical information and by a complicated presentation of test scores. In short, assessment illiterates lack the tools to be critical consumers of assessment data.

Those who are assessment literate, on the other hand, have a basic understanding of the meaning of high- and low-quality assessment and are able to apply that knowledge to various measures of student achievement. Assessment literates ask two key questions about all assessments of student achievement: What does this assessment tell students about the achievement outcomes we value? And what is likely to be the effect of this assessment on students? Assessment literates seek and use assessments that communicate clear, specific, and rich definitions of the achievement that is valued.

Just sensing a problem is not enough. Those who care about students either make or demand changes in unsound assessments.

Assessment literates know what constitutes high-quality assessment. They know the importance of using an assessment method that reflects a precisely defined achievement target. They realize the importance of sampling performance fully. They are aware of extraneous factors that can interfere with assessment results. And they know when the results are in a form that they understand and can use. Assessment literates have built-in alarms that sound when an assessment target is unclear, when an assessment method misses the target, when a sample of performance is inadequate, when extraneous factors are creeping into the data, and when the results are simply not meaningful to them.

But just sensing a problem is not enough. Those who care about students either make or demand changes in unsound assessments.

HOW ASSESSMENT LITERATE ARE WE?

To gauge our assessment literacy as a nation, we need only consider the amount of interest, time, and resources we have devoted to creating an assessment-literate society. For example, who trains students to be critical consumers of the feedback provided by their teachers? Who prepares students to act in their own best interest when it comes to the evaluation of their achievement? No one. Indeed, how many teachers are open to such critical review of their assessments?

Who teaches parents to understand the meaning of grades or of test scores that are sent home? Who trains parents to ask the probing questions of a critical consumer of data? No one. Moreover, most parents received no such help during their own school or college years.

Who helps members of school boards or others in the community to understand the meaning of high-quality assessments, and who helps them to arm their assessment alarm system? Furthermore, who takes responsibility for the assessment literacy of legislators or other public officials? No one has offered to lead these efforts. While our reports of test scores are often accompanied by interpretive guides, we rarely provide the background

training needed to go beyond the reports—to get at the real meaning of the assessment results.

When we examine the assessment literacy of those who teach and fill positions of instructional leadership, the picture becomes almost frightening. A vast majority of the 50 states require no explicit training in assessment as part of teacher certification.[1] Most states simply require completion of an accredited teacher education program. But a majority of teacher education programs require no training in assessment for graduation.[2] Furthermore, even when assessment training is offered to teachers, it typically fails to provide the kinds of knowledge and skills needed to produce assessment literates.[3] To complete the picture, those trained to become administrators are often trained less well in basic assessment than are the teachers whose work they are supposed to supervise.[4]

Is there no one who meets minimal standards of assessment literacy? Clearly, one group does: those practitioners and scholars who make up the field of educational measurement. They have the knowledge that is sufficiently refined to allow them to understand both the practical meaning and the importance of assessment. Why then has this wisdom not spread throughout the land to decision makers at all levels? If we know what it means to be assessment literate, why do so many still remain illiterate?

I cannot answer this question conclusively, but I can speculate on some of the possible reasons. One contributing factor may be that we have traditionally valued process more than product. For example, high school graduation decisions have typically been based on units completed rather than on knowledge and skills attained. Institutional accreditation standards have traditionally focused on process variables. People who place little value on outcomes are not apt to be concerned about the assessment of those outcomes. Recently, however, we have seen a growing interest in outcomes-based education, a movement that makes assessment literacy essential.

Even when assessment training is offered to teachers, it typically fails to provide the kinds of knowledge and skills needed to produce assessment literates.

Another barrier to the spread of assessment literacy may have been the fact that assessment training traditionally has been based on a very narrow definition of assessment—i.e., large-scale, standardized, paper-and-pencil test formats. This limited focus has failed historically to meet the needs of teachers. And so, over the years, those who set the agenda for teacher training have seen assessment literacy as a waste of valuable credit hours and have eliminated it from programs. There is evidence that such content insufficiency continues to characterize measurement courses and textbooks today.[5]

High school graduation decisions have typically been based on units completed rather than on knowledge and skills attained.

Others argue that courses in measurement and evaluation have traditionally set their technical standards too high for education majors and so have become unpopular and were eliminated. There is evidence that course content has traditionally included a great deal of sophisticated technical/quantitative content.

Furthermore, it may be that teachers and principals believe that systematic assessment of the achievement of their students is risky. Systematic assessment of clearly stated achievement targets may make public the fact that students already know what the school or teacher had planned to teach them or that they did not learn what the teacher or school tried to teach them. Better to keep the target and the assessment of that target vague. Then critics have more difficulty pinpointing the inadequacies of school programs. Better to be accountable for carrying out the required activities (e.g., finish the text by the end of the year) than for attaining the desired outcomes. In such an environment, assessment training is not useful.

Yet another barrier to the spread of assessment training may be the perception, because the quality of assessments is guaranteed in other ways, that teachers don't need assessment training. For instance, textbooks often come with built-in tests. Teachers can use them and so don't need to know how to construct their own tests. Furthermore, computerized item banks can provide test items and keep teachers from having to write their own. In

addition, our standardized tests are of high quality and will protect our standards even if teachers cannot. Given these sources of existing tests and test items, we might conclude that teachers are being "taken care of" and that they need no training in assessment.

What about noneducators? Why has assessment literacy eluded them? Cynics might argue that we keep the public illiterate in order to minimize scrutiny of the quality of educational outcomes or that noneducators have trusted educators to be the keepers of the standards on the naive assumption that those standards are being left in the hands of assessment literates. However, I fear that the answer is much simpler. I believe that the public—students, parents, school board members, education writers, legislators, and other private and public officials—remain illiterate with respect to assessment because too few assessment literates are available to train them.

There are as many domains of knowledge to be mastered as there are school subjects. Students need to be able to think and solve problems in each of those content domains.

In short, we in the measurement community have been poor shepherds of assessment wisdom. We have failed to find productive ways to share our knowledge with those who need it. Now, when decision makers at all levels need and want to understand and use outcome data, the cost of our failure to communicate is becoming apparent. Now we must take action to see that the long-standing embarrassment of assessment illiteracy is corrected.

ACTION 1: UNDERSTANDING ASSESSMENT LITERACY

All who presume to assess, evaluate, and act on student achievement data must come to understand the full range of possible student achievement targets and assessment methods at our disposal. Four broad categories of achievement targets are: (1) substantive *subject-matter knowledge* to be learned, (2) *thinking skills* to be demonstrated, (3) specific *desired behaviors* to be exhibited, and (4) *products* with specific attributes to be created. Each of these can be defined in many ways. There are as many domains of

knowledge to be mastered as there are school subjects. Students need to be able to think and solve problems in each of those content domains. Our expectations regarding desirable behaviors include psychomotor skills, communication skills, artistic skills, and more. And the products we expect students to create range from written papers and reports to science projects and works of art. In specific learning contexts, our achievement targets might take any combination of these forms.

The assessment-literate person can describe the achievement targets in any specific context in specific terms. He or she can describe the knowledge to be learned, the forms of thinking to be mastered, the particular behaviors to be demonstrated, and the products to be created. A person who is assessment literate realizes that effective instruction, learning, and assessment are all impossible without these specifications.

Three categories of assessment methods are available to gauge student attainment of our expectations: *paper-and-pencil* assessments, *performance* assessments (or measures based on observation and judgment), and *direct-personal communication* with the student. Each can take various forms. Paper-and-pencil methods include teacher-made tests and quizzes, tests and quizzes included in textbooks, published standardized tests, and homework and seatwork exercises and assignments. Performance assessments can focus on observations of many different behaviors or on judgments of individual products. And personal communication can take the form of student answers to questions during instruction, interviews and conversations with students, and intuitions and feelings about students.

Assessment literates know these rules and can distinguish sound from unsound uses.

Each of these methods yields legitimate achievement data when used with an appropriate achievement target in the proper context. Each has specific rules of evidence that enable one to use it well. Assessment literates know these rules and can distinguish sound from unsound uses.

To reiterate, assessment literates know if and when an assessment appropriately reflects a clearly defined

achievement target. They are sensitive to the appropriateness to the intended target of the sample of student performance being assessed. They can identify factors that may interfere with results and mislead decision makers. And they know whether or not they can use the results in the form produced. They also know that, if the appropriate conditions are not satisfied, achievement data may be rendered less useful, bad decisions may result, and students may be harmed. Assessment literates care about high-quality education and act assertively to prevent unsound assessment.

ACTION 2: DIFFERENTIATING LEVELS OF LITERACY

Clearly, the most critical action we must take if we are to improve assessment literacy nationally must be to launch much-needed programs to promote our objective. But first, we must define the scope of those programs, and that scope will differ depending on the decision makers involved. For example, students and parents clearly need not be as sophisticated in their understanding of assessment as teachers who actually construct assessments. To understand these differing needs, we must consider at least three levels of assessment literacy.

The first level is for those who are *users* of data generated by others. Such users include students, parents, school board members, legislators, most state education department staff members, education reporters, and any other private citizen or public official who uses achievement data to make decisions. These users must be functionally literate with respect to the data. That is, they must understand the decisions they make, how those decisions relate to effective teaching and learning, and what kinds of data can inform their decision making. They must also be generally familiar with the four key attributes of good data (a clear target, an appropriate sample, known sources of interference, and usable results) as they relate to each basic type of assessment. Then and only then can they ask the questions necessary to promote sound assessment.

The second level of assessment literacy is for the practically literate, those who both *generate* and *use* data.

That group includes teachers (at all levels), principals, guidance counselors, school psychologists, superintendents, teacher trainers, and trainers of administrators. These professional educators must attain a higher, more practical level of assessment literacy. Like the functionally literate users, they need to be clear about the decisions they make, about how those decisions affect students, and about what kinds of data will help. However, they also need to attain a high level of understanding of the attributes of sound data, factors that can reduce the quality of data, and specific strategies for avoiding those pitfalls. They need to know that achievement targets that we don't understand are impossible to assess, that biased samples of performance result in useless data, that interference factors must be controlled, and that results not understood by users lead to the waste of valuable assessment resources.

Professional educators need to attain a high level of understanding of the attributes of sound data.

Furthermore, unlike the functionally literate, the practically literate must be capable of developing and using all three basic types of assessment to generate quality information. They must be able to construct paper-and-pencil instruments to assess knowledge and thinking skills. They must be able to design assessments, based on observation and judgment, that focus on important behaviors and products. And they must be able to interact directly with students in ways that give insight into student learning. They need to know and understand the strengths and limitations of each assessment option. They must also be able to defend the soundness of their assessments.

The third level of assessment literacy is for advanced literates, those who *generate data for use by others.* They prepare the "high stakes" tests—tests used to certify minimum competence, to select students for special treatment, to survey achievement on a wide scale, and to evaluate programs of instruction. These individuals include all specialists in measurement and evaluation—those who must apply an advanced level of sophistication in addressing assessment issues. These professionals must master all that is required for functional and practical

literacy. But they must also master large-scale test development, administration and scoring, and classical and modern psychometric theory, as well as having the experience to create large-scale assessments of demonstrated validity, reliability, and economy.

Assessment literacy is a multidimensional concept. As the context varies—both in terms of the needs of the decision maker and the practical realities of the decision context—the meaning of assessment literacy varies. However, all meanings stem from a clear sense of the decisions to be made and a fundamental understanding of the attributes of sound data. All data users must become critical consumers.

But those who actually select and develop assessments must be held to a higher standard. Teachers and those who train them, as well as all administrators and those who train them, must understand the full range of assessment methods at their disposal.

ACTION 3: ASSESSMENT LITERACY PROGRAMS

We expend huge amounts of money in pursuit of school improvement and valued educational outcomes in a society in which educators and noneducators alike can neither define nor systematically assess those outcomes. This is unacceptable. What will it take to create a society that is assessment literate? We must start with the right attitudes and then undertake a great deal of basic assessment training. Everyone must contribute if we are to succeed.

Let's begin with the educational measurement community—the reservoir of current expertise. Members of this community (and I am one) must adopt the attitude that teachers *do care* about outcomes and do want to learn how to assess them fairly, thoroughly, and dependably. In fact, they do care—very much. Furthermore, we must realize that large-scale assessments represent only a small fraction of all assessments conducted in schools. We must acknowledge that the assessments that count in terms of student learning and academic self-concept are those used every day in classrooms. Large-scale assessments have been the focus of our research and development, and, as a result, their quality has improved markedly over the years.

Now the time has come to refocus our research and development resources and efforts on classroom assessments.

As we do so, we in the measurement community must assume our share of responsibility for the lack of assessment literacy, and we must take the lead in developing new training efforts to rectify the situation. We must find creative ways to demystify the concepts and practices of technical measurement. We must further refine our understanding of various levels of assessment literacy and begin to deliver critically needed training at all levels. But simply offering sound, relevant college and university measurement courses and writing helpful measurement textbooks, while very important, will not suffice. It is essential that we provide high-quality, efficient inservice training for all those teachers and administrators who have completed their degree programs without relevant assessment training.

Furthermore, because assessment literacy is multidimensional and varies in meaning depending on level, the training should be focused on the particular needs of the user. That means that, for those who need functional or practical literacy, some skills (e.g., labeling and estimating various forms of reliability and validity or equating test scores) are not nearly as important as others (e.g., specifying achievement targets and developing practical measures that match them). It may be that many users have shied away from assessment training because they fear—perhaps for good reason—that the training will be more technical and quantitative than they need or want it to be.

Only when the ultimate consumers of data press for real clarity will educators become assessment literate enough to provide clear explanations.

How can teachers and educational administrators contribute to the effort to improve our assessment literacy? The time has come to face the realities revealed by sound assessment practices and to accept responsibility for student outcomes—when they are measured well. High-quality, classroom-relevant training in assessment should have been part of the professional preparation all teachers and administrators paid for. It was not. They should

demand it now. If and when such training is made available, they should take advantage of it to become assertive, critical consumers of assessment data. Then they should keep their alarm systems armed and ready. When a new assessment sets off the alarm, they should begin to ask critical questions and not stop until they are sure that students are well-served by assessment—whether in the classroom or at other levels of decision making.

How can students and parents, school board members, legislators, and members of the news media help in the effort to make this an assessment-literate society? They must assert their rights and leadership and become literate themselves. Students and parents must resolve not to be intimidated by educators. They must tune their alarm systems to the potential problems of clarity in assessment targets, of adequacy of samples, of factors that can interfere with assessments, and of clarity in presenting results. They must act in their own best interest to point out unsound assessments. Policy makers must satisfy themselves that they understand the meaning and quality of outcome data. They must continue to demand explanations of test data until they are sure that they understand. Only when the ultimate consumers of data press for real clarity will educators become assessment literate enough to provide clear explanations.

THE BOTTOM LINE

One need not search far to find what appears to be compelling evidence that, as a society, we really don't care whether students learn anything in school or not. Consider this evidence.

To begin with, if we really cared about student learning, would we not have focused our research and development efforts over the decades on understanding effective ways to assess that learning both in the classroom (where almost all school assessment occurs) and in the large-scale assessments? In fact, we have conducted very little research on classroom assessment, devoting nearly every research and development dollar to the improvement of standardized testing. As a result, as we enter the 1990s, we still do not understand the critical task demands

of classroom assessment.[6] Does the narrow focus of our assessment research reflect a profession that truly values achievement?

If we really cared about student achievement, would we not go to great lengths to be sure that every teacher was able to assess achievement in the classroom every day? In fact, we rarely offer teachers the assessment training needed to do so. Nancy Conklin and I studied the assessment training requirements of the 27 undergraduate and graduate teacher training programs that produce 75% of all teachers trained in the Pacific Northwest. We could find only six required measurement courses.[7] Furthermore, students who *are* trained are still inadequately prepared with respect to assessment concepts and procedures needed to address the ongoing assessment demands of the classroom.[8] When so few teachers are trained to assess student learning, does this sound like a profession that cares whether learning is occurring?

As we enter the 1990s, we still do not understand the critical task demands of classroom assessment.

When we examine the professional preparation of administrators with respect to the assessment of student achievement, we see an even more troubling picture. The vast majority of administrator training and certification programs offer no training in assessment at all.[9] Many simply cannot evaluate assessments and therefore do not know whether sound assessments are being used in the classrooms in their schools. Would a profession that cares about learning not prepare its leaders to verify that learning is occurring?

There is more. Would a profession that really cared about student learning not impose standards for high-quality assessment on those who develop the tests that accompany published textbooks and other curriculum materials? Such standards exist for standardized tests, but no such standards have been advanced for the tests and quizzes embedded in textbooks. The effect is often apparent in the low quality of these assessments.

Finally, would a profession that really cared about outcomes not have found a way to systematically train its policy makers (school board members, state department

personnel, state legislators, federal policy makers, and so on) to understand and use assessment data in an informed and appropriate manner? In fact, such understanding is rare indeed.

Despite all this evidence, my point is not that we don't care about learning. Of course we do. My point is that we spend all of our available resources to train teachers and administrators to produce learning and to put that training to work in schools. Then we allocate virtually no resources to train practitioners or interested others in sound methods of assessing the outcomes of those efforts.

Students count on their teachers to offer sound instruction and sound assessment. Sound instruction involves sound decision making—at all levels. And sound decisions require sound data. When assessment-literate decision makers understand the differences between sound and unsound data, they can provide the effective and efficient instruction that students count on. When assessment illiterates fail to understand and detect unsound data, the students suffer most. The time has come to promote assessment literacy for all.

NOTES

1. Peter Wolmut, "On the Matter of Testing Misinformation," paper presented at the SRA, Inc., Invitational Conference, Phoenix, 1988.
2. William D. Schafer and Robert W. Lissitz, "Measurement Training for School Personnel: Recommendations and Reality," *Journal of Teacher Education*, vol. 38, 1987, pp. 57-63; and Richard J. Stiggins, Nancy F. Conklin, and Associates, *In Teachers' Hands: Investigating the Practice of Classroom Assessment* (Albany, N.Y.: State University of New York Press, forthcoming).
3. Arlen R. Gullickson and Kenneth D. Hopkins, "Perspectives on Educational Measurement Instruction for Preservice Teachers," *Educational Measurement: Issues and Practice*, vol. 6, 1987, pp. 12-16; and Stiggins, Conklin, and Associates, op. cit.
4. Schafer and Lissitz, op. cit.
5. Stiggins, Conklin, and Associates, op. cit.
6. Richard J. Stiggins, Nancy F. Conklin, and Nancy J. Bridgeford, "Classroom Assessment: A Key to Effective Education," *Educational Measurement: Issues and Practice*, vol. 5, 1986, pp. 5-17; and Gary Natriello, "The Impact of Evaluation Processes on Students," *Educational Psychologist*, vol. 22, 1987, pp. 155-75.
7. Stiggins, Conklin, and Associates, op. cit.
8. Gullickson and Hopkins, op. cit.; and Stiggins, Conklin, and Associates, op. cit.
9. Schafer and Lissitz, op. cit.

Performance Assessment: The Message from England

by Desmond L. Nuttall

Although many educators in the United Kingdom think performance assessment benefits students, critics are challenging the rigor of the evaluations and advocating dropping practical tasks from examinations.

All the world has become interested in performance assessment. It is now agreed, even in the U.S.A., the home of multiple-choice tests, that such tests do not measure all the important aspects of learning, especially higher-order thinking and practical skills. In England, where performance assessment is not new, there is little doubt that this form of evaluation can improve teaching and learning. Nevertheless, many practical tasks are being dropped from national examinations to simplify the process and save money. Let's look back in time to understand why the trend toward authentic forms of assessment went into reverse.

THE ASSESSMENT OF PERFORMANCE UNIT

England's Assessment of Performance Unit (APU), established in the mid-1970s, fulfilled a similar function in England to that of the National Assessment of Educational Progress in the United States: it provided a light monitoring of educational attainment over time. Early NAEP surveys, like some early state surveys (for example, Michigan's), assessed a wide range of tasks, including practical activities, but subsequently dropped them mainly on financial grounds.

The history of the APU has been somewhat different. The surveys used two methods: traditional assessment and

From *Educational Leadership*, vol. 49, no. 8, p. 54-57, 1992. Reprinted with permission.

practical assessment. The practical surveys, which are very imaginative, especially in science and mathematics, required the assistance of visiting assessors to test children one at a time, or in pairs. Although many had feared that the sampling of the APU would rapidly evolve into a method of comparing the performance of different school districts or even of different schools, in practice, teachers welcomed it. They valued the work of the APU for its twin effects of describing student achievement on a much broader front than was typically tested and for providing a mine of ideas for classroom assessment.

PUBLIC EXAMINATIONS

Meanwhile, the traditional public examinations taken by most of the population in Grade 11 and by the elite (around 20 percent of the population) in Grade 13, remained in much the same old format. At Grade 11, there were two examinations, one for the highest attaining students and the other for the rest of the population (mirroring the old division between selective grammar schools and the rest). As comprehensive schooling gained ground, the dual examination system lost its basic rationale, and by 1970 there were calls for a merger.

After a major program of feasibility and development studies, Her Majesty's Inspectorate of Schools, a group with high prestige and much influence in education, recommended the establishment of a common system of examining, including more performance assessment. The Inspectorate's reasons are not hard to find. In a national survey of secondary schooling, published in 1979, the group made a damning indictment of teaching and learning in the two years leading up to the old Grade 11 examinations:

> ...the effect of the dominating pursuit of examination results was to narrow learning opportunities.... In at least one-fifth of the schools the demands of public examinations appeared to be an important factor in the impoverishment of reading.... In many subjects writing tended to be stereotyped and voluminous.... In mathematics, examinations were one factor in encouraging narrow, repetitive

> practice and standard routines divorced from applications to real situations.... Similarly in science, the emphasis on content and concepts unrelated to the pupil's experience... was precluding investigational activity even where examinations required assessment of such skills.

The Government, never hasty in implementing other people's ideas, took eight years to agree to establish a new examination, the General Certificate of Secondary Education (GCSE).

Every subject in the GCSE contains some element of performance assessment: investigational work in mathematics, oral work in English and in modern languages, portfolios in English, and projects (extended writing up of an investigation carried out and usually chosen by the candidate) mostly in the humanities and the social sciences. Most of this work is graded by the candidate's own teacher, subject to monitoring by an examining body. Monitoring usually occurs through visits by a part-time moderator whose own grading standards are extensively monitored by a Chief Moderator (who is also responsible for training the moderators).

Every subject in the GCSE contains some element of performance assessment.

In the four years that the GCSE examination has been running, the proportion of students awarded the top three grades (of the seven that can be achieved) has increased substantially. Most people hail this as proof that educational standards are rising (even allowing for some grade inflation as a result of greater familiarity with the examination on the part of teachers, students, and parents). Some, though, view the increases as evidence that examiners' standards are slipping and that performance assessment is making the examination far too easy. They point out that students (especially middle-class ones) can get considerable help with assignments from their parents and that essays can be polished before they become part of the portfolio.

The majority's view, however, is that GCSE has been an educational success. Learning and teaching have changed to accommodate the new forms of assessment, and there are now many more student-centered activities

in Grade 10 and 11 classrooms. Assessment can, it appears, be a device to change (and improve) teaching and learning. This lesson has not gone unnoticed in the United States. Michael Cohen and Lauren Resnick, who direct the New Standards Project, are experimenting with ideas for a national examination system based heavily on performance, which would help to motivate young people to learn and foster a broader range of skills in U.S. classrooms.

NATIONAL CURRICULUM ASSESSMENT

The power of assessment to control the curriculum and what goes on in classrooms, and perhaps to gain votes, did not escape the Thatcher Government. Its 1988 Educational Reform Act was one of the most substantial pieces of English legislation on education this century. Among many other things, the act mandated the creation of a National Curriculum and an associated assessment system for all students in Grades 2, 6, 9, and 11 (in the last case incorporating GCSE). The National Curriculum was clearly designed to place what was taught in the schools under the control of the Government.

Officials, however, paid less attention to how the curriculum was taught and assessed, leaving the design of an assessment system to a panel of educationists. The panel followed current good practice, as exemplified in the GCSE examination, and proposed extensive performance assessment. The system was to be an amalgam of the teacher's assessment of the progress of each child (against national standards as defined in the National Curriculum) and the child's performance in a series of Standard Assessment Tasks designed centrally but administered and graded by the child's own teacher. The activities were to consist of ones very like those employed every day in the classroom, so much so that the children might not even realize that they were being assessed; the "Standard" in "Standard Assessment Task" simply indicates that the task is common across England. Small-scale trials were held in

Standard Assessment Tasks were to consist of activities so very like those employed every day in the classroom that the children might not even realize they were being assessed.

1990, and a fully fledged national pilot (for every Grade 2 child in the kingdom) was held in the summer of 1991.

One could hardly recognize the 1991 Standard Assessment Task as a test. First, it was sent out weeks before the assessment period to enable teachers to dovetail it into their own teaching plans. Second, no standardized instructions were prescribed; teachers were encouraged to introduce the tasks in their own words. They were to be done either individually or in small groups over a period of at least three weeks and were estimated to occupy about 30 hours of teacher time.

EXAMPLES OF STANDARD ASSESSMENT TASKS

Examples from English and science illustrate the range of tasks assessed.[1]

English. Teachers are to select stories targeted at the average 7-year-old "that will give rise to an extended discussion." Additional instructions to teachers include:

- Read the story aloud to a small group of children.
- Lead a discussion about the story; encourage children to contribute fully and to think further about their contributions.
- Ask further questions, if necessary, to draw out children's understanding of the structure of the story.
- To be successful in this assessment, children need to understand story structure in terms of the opening, the unfolding events, and the conclusion. Their contributions should be more than just short answers to direct questions.

Science. This task, which is aimed at the average 7-year-old, starts with a long list of materials the teacher might need, including containers of water for floating and sinking investigations. The instructions to teachers continue:

- Set out all the apparatus and give each child four objects including: (1) at least one object that has some features which are easier to see using a hand lens (for example, a twig); (2) at least two objects that will float (for example, a twig and an apple); (3) at least one object that will sink (for example, a stone); (4) at least one heavy object that will float (for example, a piece of fruit).

- Present this part of the activity to the children as a challenge to find out as much as they can about why some objects float and some sink.
- Ask the children to tell you *what* they think will happen if one of their objects, which you name, is placed in a container of water—and *why*. With each child, repeat this with a second object.
- Explain to the children they are to draw, weigh, make a prediction whether each object will float or sink, and then test each object to see whether it will float or sink. On the charts supplied, have them record their predictions, the outcomes of their tests, and the weight of each object.
- Explain to the children that they should observe and begin to draw each object without using a hand lens and then finish the detail of their drawing of each object with the use of a hand lens.
- Allow the children to carry out their practical work and complete their charts and drawings.
- When the children have completed their activity, ask each child what he or she found out about why some objects floated and some sank.
- Now ask each child three questions to see if he or she can collate information from the chart.

If children meet the standards for the average level, teachers are instructed to have them proceed to the next level (again defined by explicit standards). The next task is for the children to investigate the hypothesis that heavy objects sink.

EVALUATION OF THE ASSESSMENT

In most cases, the children enjoyed the activities on the 1991 Standard Assessment Tasks and were not made anxious by them. Many teachers said that they had learned new things about the children's attainments, adding that they had often not rated the children as high as the assessment did. However, they found the tasks to be demanding of their time and energies; invariably, to prepare, administer, and grade them required an average of 44 hours. In addition, it was almost impossible to administer the Standard Assessment Tasks and look after the rest of the

class simultaneously. Most teachers were able to arrange for another teacher to assist with the class, but the interruption of normal education was substantial, as were the extra costs incurred.

Also, as noted earlier, some people have felt that performance assessment is making the examination less rigorous. Unfortunately, the minority who take this view have the ear of Prime Minister John Major, who has decided that the amount of performance assessment must be significantly curtailed. His exact words were: "By testing I do not mean some weird experiment in a corner. [A reference to the Standard Assessment Tasks.] What I mean is pencil-and-paper testing for a classroom so that people have a measure of how they are doing—see if there is a problem so that you can put it right."

Performance assessment can be made to work, and both teachers and students welcome it.

As a consequence, the plans for national assessment at Grade 2 in 1992 (and in later years for the other grades) have been changed dramatically. For example, the day after the Prime Minister's statement, the development contracts for Grade 9 Standard Assessment Tasks were canceled, and new tenders were invited for the creation of paper-and-pencil tests to the virtual exclusion of the sort of practical and naturalistic tasks that had been under development. At Grade 2, also, the tasks have been redesigned so that many can be administered to the whole class at the same time, and some of the most time-consuming ones have been dropped. In 1992, the administration of the Standard Assessment Tasks should be much simpler and take up much less time. In addition, teachers will have had ample time to prepare for it, since it was circulated several months before it had to be given.

REPORTING OF RESULTS

One aspect of National Curriculum Assessment that has not yet been put to the test is the reporting of results. A major justification for doing so is to give parents more information about their child's progress, but there are other agendas.

Scores will be made public school by school and classroom by classroom. The publication of the former in local league tables is designed to give information to parents that will help them in choosing a school for their child. The publication of the latter risks a return to the Victorian era, when teachers were paid according to the attainment of their students.

Publication of results in this manner puts much emphasis on comparability of grading between different schools and a move to more objectively scorable tests. The introduction of a more rigorous system of moderation will surely be demanded when scores begin to be published (probably in 1993). Making fair comparisons also requires comparing like schools with like (and like classrooms with like), as is done in California and several other states. The Government is resisting anything but the publication of "raw results," but it is almost certain that school districts will also produce results taking the characteristics of the student body into account.

LESSONS FROM ENGLAND

Performance assessment can be made to work, and both teachers and students welcome it. Through the new GCSE examination in England, the process has improved teaching and broadened the range of skills appraised. Performance assessment demands new skills of the teacher as well, and a new professionalism. Performance assessment can only take off with the will and expertise of the whole profession and with the trust of parents and politicians.

I suggest that the United States might learn two other things from the English experience: first, the cost of performance assessment, both financially and in terms of the time of teachers, is immense; and, second, despite all the care and effort, some will still not view it as rigorous enough.

NOTES

1. Source. *1991 Key Stage 1 Teacher's Book*, (London: Department of Education and Science/Welsh Office).

But...the Curriculum

by Carolyn Mamchur

Ms. Mamchur poses 10 tough questions for secondary teachers who bemoan their students' lack of interest, their laziness, their resistance to learning.

JUNE SIKLER was tired. She had been marking papers half the night. So her patience wore a little thin when her 11th-grade English class responded to her request to "go over the errors and rework your papers" with groans, quick sidelong glances, and a general lack of interest. "Okay, fail then," she felt like saying. "You'll have to stay after class and write out each story five times," she thought of threatening. "Don't come to my class if you don't want to work. Take a little visit to the principal's office. Just get lost," she wanted to say. But she didn't. She made her voice even and low, she smiled a thin smile, and she walked between the rows and commanded obedience with her small, tight body.

In all my 25 years of teaching...I have *never* known teachers who are having trouble with discipline or motivation to say that their students have choices, control, or true responsibility for their schooling.

Like everyone else, secondary teachers get tired and frustrated—even fed up—from time to time. They see their best-laid plans and hardest work wasted on students who seem not to care. Their frustration is to be expected. It's like mild depression or occasional migraine—it comes, it goes. Usually it comes when a teacher is tense, overworked, and in a stressful situation.

Although it may seem so at the time, it is not always the students who put the teacher in such a situation. If

From *Phi Delta Kappan*, vol. 71, no. 8, p. 634-637, April 1990.
Reprinted with permission.

teachers tend to blame students—to bemoan their lack of interest, their laziness, their resistance—then perhaps the teachers had better ask themselves a few tough questions.

When things were going wrong for me as a secondary teacher, I asked myself 10 tough questions. I now ask them of my preservice students when they are having difficulty in their teaching practicum. The same 10 questions keep popping into my mind when teachers or school districts request inservice workshops on motivation or discipline or classroom management. For 25 years now—as a high school teacher, as a teacher educator, as a trouble-shooter—I have been asking the same set of questions. Either they are awfully important questions, or my mind is stuck on them —imprinted, like the images on a computer screen left on bright when the owner goes to the south of France for a three-month vacation.

These are the questions imprinted on my mental screen:

1. **How often do students have choices in the classroom?**
2. **How often do students feel in control, in charge of themselves?**
3. **How many decisions that really count are students allowed to make?**
4. **How are students responsible for their own learning?**
5. **How do your students know when they have pleased you?**
6. **How do you know when you have pleased them?**
7. **How often is there a right answer to the questions you ask of your students?**
8. **How often do students feel important in your classroom?**
9. **How often do you and your students share laughter and pleasure?**
10. **How do the students react when they walk into your classroom? Happy? Calm? Safe? Excited? Assured? Afraid? Bored? Sorry?**

In all my 25 years of teaching secondary students and training teachers, I have *never* known teachers who are

having trouble with discipline or motivation to say that their students have choices, control, or true responsibility for their schooling. Yet when I ask why students don't have control, I hear a standard set of responses. "We have no time to allow for choices—the curriculum has to be covered!" argues one teacher. "Students aren't able to make decisions regarding curriculum," asserts another. "It's on the exam, and that's that!" a third states flatly. "This is secondary school!"

An analysis of the questions themselves reveals that they deal with power, not with curriculum. Could it be that secondary teachers hide behind the safe excuse, the cop-out excuse, the it's-always-been-this-way excuse, the excuse that blames curricular requirements when what we really want is power? "I believe that the need for power is the core—the absolute core—of almost all school problems," William Glasser told *Kappan* readers.[1] Could it be that, if we gave power to our students, they would cease to appear lazy, uncaring, and undisciplined?

Does giving students "power" mean choosing "students" over "curriculum?" Not at all. Curriculum is the tool.

Teachers who have not tried to run their classrooms in ways that allow students to satisfy their need for power often mistakenly worry that in giving power to students they will lose their own. Teachers who encourage their students to make choices, to take responsibility for their learning, to be independent and fulfilled in classrooms in which autonomy, decisiveness, risk-taking, and caring are recognized and appreciated know that power shared is not lost—but increased. In a power struggle, on the other hand, someone always loses.

* * *

Does this mean that the teacher must stop being concerned about the curriculum? Does giving students "power" mean choosing "students" over "curriculum?" Not at all. Curriculum is the tool, the "stuff" of education. Through curriculum students learn the skills, the attitudes, and the knowledge they need. Curriculum is a vehicle, a device.

As an English teacher working with "tough" students, I was asked by the school board to leave classroom teaching and become a counselor. "Why?" I asked.

"Because students don't have discipline problems or truancy problems or 'failing exam' problems with you. You're good with them."

"*I'm* not good. The way we work together is good."

"Well, work with more students, as a counselor."

Work with what? I asked myself. The "stuff" we worked with in my English classes was the curriculum. Literature, drama, writing. How could I teach without these? *I* didn't keep my students quiet, interested, working. They did that themselves, using curriculum they cared about in ways that were interesting and meaningful to them. They gained power over language, thought, and ideas. And they had so many choices to make: What stories will we study? How should we choose? How can we express our ideas about the stories? How can we help one another? What use can we make of our new knowledge? They read, wrote, put on plays, made films, bought an old school bus, went on the road, entertained in nursing homes, and tutored other young people having academic problems. How could you keep these students *out* of school? Have you ever visited the home of a really "troubled" youngster? Did you see much peace and quiet? Much time or energy devoted to caring? Much excitement over what the child had to say or was doing? Troubled young people come to school starved.

Curriculum must never be the "stuff" we rub our students' noses in—as if they had just dirtied the carpet.

Curriculum can become a delicious feast, tempting students to taste, to satisfy their appetites. Having once eaten, though, they are not forever full. Indeed, tomorrow's hunger will be even keener, and the palate will be even more sophisticated. Curriculum must never be the "stuff" we rub our students' noses in as if they had just dirtied the carpet; that doesn't do much for the appetite.

The happy truth is that students learn the curriculum best when we give them power over how they learn it. But what about students who don't seem to want to cooperate? We can be sure that such students won't learn the curricu-

lum by force. What have they—or we—got to lose by trying an alternative method? Forced smiles, thin voices, tight bodies? Which method is easier on us, their teachers? Which one makes us feel good about ourselves at the end of the day, the week, the year?

Giving students choices may seem to be a complex issue. But actually, it is dead simple. The rule is this: whenever you can give a student a choice of any kind, do it.

Some of these choices will be big ones. Early in September a young teacher working in a remarkably student-centered school phoned me for some advice. She and two other teachers were dividing fifth- and sixth-grade students into three groups to work on a science project. She knew that I had done work on learning styles, and she wanted to know if it would be wise to put these students into groups matched for learning style.

"How are you organizing the project?" I wanted to know.

"We are having three themes: the human factor, research, and the world around us."

"Why not let each student choose the area of study that he or she wants to start with? Try to help the students make informed choices. Describe what's going to happen in each area, and suggest that they base their choices on their own interests and abilities."

"Won't they have to take part in all three? We planned for them to do all three," she said. She sounded worried.

I assured her that the students would eventually work in all three areas. But why not let them choose the order? It makes sense, and research bears out the fact that "matching" students' learning styles, interests, and abilities to a task at the introductory stage is effective.[2] "Mismatching"—or moving to the more difficult or unfamiliar tasks—should take place once students have built up confidence and a more sophisticated level of interest and knowledge.

"What about learning styles?" the teacher wanted to know. "Shouldn't we group that way at all?"

The power of understanding differences in students' learning styles is that it allows the teacher to offer better-

informed choices. A teacher can design learning centers, assignments, or tests in a variety of ways. By designing them with the different learning styles of the students in mind, the teacher can feel confident that all students will have the opportunity to learn in a way that suits them best.

Matching assignments to learning styles works even better if the students are also informed. The teacher could help students to understand their own learning-style preferences; then, when they make choices—in projects, in ways to learn new concepts, or in ways to demonstrate mastery—or when they find themselves performing better in one mode than another, the students will understand why. The students might even develop their own alternatives. The one thing better than having students choose from a variety of modes of learning is having them generate their own options.

* * *

Choice can come in many shapes. For example, there is the large choice of which programs to begin studying. Last spring a friend of mine was challenged by a school district to conduct a workshop on "the writing process" for resistant learners. She accepted the challenge on the condition that she could offer the students choices throughout the three-day workshop. The first choice she offered her group of 78 students from grades 7 through 11 was this: "If you would, for any reason, rather not be in this workshop, please don't stay. You are free to go to the library, to your regular class, or to the sunny side of the building. The one choice you don't have is to stay in this workshop if you don't feel that this is a good time to learn to improve your writing skills."

That was the big choice. At the beginning of each session she repeated the offer. By the end of the three days, she had lost 22 students. Of the 56 who remained, every one completed a polished composition—many, for the first time that school year.

Every chance she got, my friend offered these students choices. They were allowed to choose the topics they were going to write about. She suggested the general topic of "an awkward moment." She suggested subtopics related to being happy or surprised or embarrassed or shy or

frightened. She used an overhead projector to share several of her own pieces of writing. She let students choose the ones they liked best and asked them to tell her why. She invited the class to ask her any questions they wished about her awkward moments or about the ways she chose to write about them. She offered students the opportunity to write about their own awkward moments, to write on any other topic, or to act as observers/reporters for the next hour—recording what was happening in the class in order to explain it to their teachers when they went back to their regular classrooms.

The students were free to determine the genre, the length, and whether they wanted to talk over their intended topics with partners before writing or to read more samples of awkward moments written by the teacher or by other writers. They chose whether to speak into a tape recorder or to write their papers—and, if they chose the latter, whether to use a pencil or a pen or a crayon or a computer. They could work in pairs, in threes, or alone. They could sit on the floor, at a desk, or at a table. They could listen to music with headphones while they wrote. They could *not* choose to interfere with the work of others or to have no task.

If you give choices, you must mean it. Our job is not to judge students' choices but to work with them.

Teachers marveled at the success my friend was having. One who had taken a previous workshop with her asked permission to sit in on this workshop during the second day. She had tried my friend's methods of teaching the writing process with limited success.

During the second day's session a young man who was considered severely learning disabled had difficulty understanding the idea of "showing, not telling." He simply couldn't grasp the concept. My friend, who is very open and fun-loving by nature, asked the young man whether he liked her. The boy grinned and shyly dropped his head. "I'll take that for a yes," she said and gave the boy's hair a gentle tousle. "How would you let me know you liked me if you were too shy to tell me?" The boy looked up at her and smiled. At recess he went to the

schoolyard and picked my friend a bouquet of daisies and put them on her desk. She went to the board and wrote the sentence, "He liked her." Then she wrote beneath it, "He brought her a bouquet of daisies the color of sunshine." She announced to the class that Ronald had just provided a perfect example of "showing, not telling." The first sentence told, the second showed.

Ronald was so proud and so confident that he even decided to try to write about his own awkward moment. His teacher, who wished to be more successful in teaching the writing process, peered over his shoulder to see what he was writing. Ronald had written, "I passed a fart in class." His teacher said, "Be serious, Ronald. It isn't nice to make fun when Miss Simpson is working with us. Don't write something stupid."

Ronald decided not to come to the third workshop session. And Miss Simpson knew why her method of teaching the writing process didn't work for this teacher. If you give choices, you must mean it. Our job is not to judge students' choices but to work with them. If we choose not to, we have made an important choice.

* * *

June Sikler is having a conference with a student. They are analyzing ways to improve the student's ability to clarify ideas. "What is your main thought here?" the teacher asks. "What did you really want to say?"

"I really believe that responsibility is a loaded word. I want to show how my ideas about it have changed," the student replies. The teacher is relaxed. She is listening carefully. The student is deeply engaged, concentrating her energy.

Miss Sikler can focus so much attention on one student because the rest of the students are also engaged. Working alone, in pairs, or in larger groups, they are busy with writing topics that they have chosen themselves. They are writing about things they know and care about. They help one another with their work, sharing ideas and skills. Their knowledge and judgment count. What they say to one another is important.

"Does my story make sense to you?"

"Can you visualize this part?"

"What does this remind you of?"

"What's another word for *lazy*?"

I wasn't there to hear the response to this last question. But when the word is applied to students, I might suggest "controlled, bored, unsatisfied, powerless."

NOTES

1. Pauline B. Gough, "The Key to Improving Schools: An Interview with William Glasser," *Phi Delta Kappan*, May 1987, pp. 656-63.
2. Daniel Solomon and Arthur J. Kendall, *Final Report: Individual Characteristics and Children's Performance in Varied Educational Settings* (Rockville, Md.: Spencer Foundation Project, 1976).

Linking Restructuring to Authentic Student Achievement

by Fred M. Newmann

Unless plans for new structures are guided by a vision of educational outcomes that articulates new standards for student achievement and for teaching, restructuring is not apt to produce favorable results for students, Mr. Newmann warns.

THE CALL to restructure education sends an ambiguous but serious message. The word *restructure* suggests that the system is so fundamentally flawed that such terms as *innovation* and *reform* aren't robust enough to describe the changes that are needed. But what exactly is the challenge that we face in restructuring? What needs to be changed, in what ways, and for what purposes?

The history of education in America has been punctuated repeatedly with powerful slogans that mobilize the energy of practitioners, researchers, and policy makers. Recent examples include "back to basics," "relevance," "community control," "cultural literacy," and "effective schools." Much of a slogan's appeal rests in its capacity to conjure up multiple meanings that can draw diverse constituencies together in a common cause. While a slogan galvanizes public opinion and focuses energy, thus offering new possibilities for action, its ambiguity brings the great risk that energy will be dissipated in scattered, even contradictory, directions.

Before committing ourselves to new structures, we should try hard to reach agreement on the nature of the system's failure.

Therefore, before committing ourselves to new structures, we should try hard to reach agreement on the

From *Phi Delta Kappan*, vol. 73, no. 6, p. 458-463, February 1991. Reprinted with permission.

nature of the system's failure. For without agreement on the nature of the illness and on the purpose of treatment, the pursuit of radical cures could do more harm than good for students.

Today, reformers tend to address two main issues. The most obvious is the large proportion of students, especially low-income students of color, who fail in school and score poorly on nationally standardized tests. Much of the rhetoric about national education goals reflects concern about this group. Another issue—less popular, but perhaps even more disturbing—is that even those who succeed in school and score well on conventional tests have not been educated to cope successfully with the demands of personal, vocational, and civic life in contemporary society.[1]

We must recognize the unfortunate record of traditional efforts to teach a virtually limitless list of fragmented pieces of declarative knowledge.

Is the point of restructuring to provide a better way of teaching the current curriculum to students who haven't learned it through the old structure? Or is restructuring to be a vehicle to fundamentally change what is taught and how it is taught to all students? These alternative aims for restructuring can reflect conflicting goals of education, and the goals held by those pursuing each aim may lead to the creation of very different types of educational environments. It may not be necessary to chose between these stark alternatives, but the tension must be recognized. It reflects a persistent historical debate about educational ends that has no simple resolution and that even the restructuring movement will not solve once and for all.

Regardless of one's position on the forms of competence that schools should promote and regardless of the fact that large proportions of parents and educators perceive the schools to be working well, there are at least two grounds for seriously entertaining the prospect of fundamental organizational change. First, teachers face the persistent difficulty of engaging students in serious academic work in schools as we know them. Except for a few highly motivated students, most young people complete

school only as a ritual. This persuasive disengagement creates massive problems of crowd control for educators and wastes the time of students and staff members alike.

Second, we must recognize the unfortunate record of traditional efforts to teach a virtually limitless list of fragmented pieces of declarative knowledge—e.g., that the issue of slavery divided the U.S. in the 19th century or that squares have equal sides and 90-degree angles. This enterprise, pursued over the ages as the main task of formal schooling, now actually undermines education. When schools try simultaneously to keep up with the explosion of knowledge and to accommodate the host of social demands placed on them, they cannot possibly give students the opportunity to develop in-depth understanding and the capacity for higher-order thinking that are required for success beyond school. The problems of disengagement and of excessive emphasis on coverage are related, and they exact their most tragic toll on students who, because of low income, cultural background, or lack of social support, do not succeed in schools as they are currently organized.

One of the biggest mistakes we make is the order in which we design the components of the education system. First, we set up an organizational structure; next, we plan the curriculum to fit the structure; and finally, we choose criteria for student success. Not until this last step do we articulate in a concrete way the ultimate educational purpose of the system. But by this point the outcomes are largely predetermined by the organizational structures and curriculum. Thus it should not be surprising that schooling seems so dysfunctional.

Moreover, the restructuring movement is about to perpetuate this error, because it, too, is trying to design organizational structures before clarifying purposes and reaching consensus on the educational ends that organizational structures should serve. For example, most of the talk one hears about school-site management, career ladders for teachers, or schools of choice never mentions how these mechanisms will teach students to write about literature, to reason about scientific phenomena, or to learn important geographic facts.

There is no reason to believe that restructuring will produce favorable results for students unless plans for new structures are guided by a vision of educational outcomes that articulates new standards for student achievement and for teaching. In this article I hope to nudge restructuring in this direction. First, I propose *authentic* achievement as the primary goal for students. Then, I advocate the need for *substantive conversation* in teaching, a new standard that is necessary to promote authentic student achievement. Finally, I suggest particular structures for student work that can facilitate authentic achievement and substantive conversation.

AUTHENTIC ACHIEVEMENT

The kinds of skills required to earn school credits, good grades, and high scores on typical tests are often considered trivial, meaningless, and contrived—by both students and adults. In contrast, a "restructured" vision of the goals of education seeks to evaluate performance activities that are worthwhile, significant, and meaningful: in short, activities that are authentic.

What criteria help us to recognize authentic forms of academic achievement? Consider the achievements of successful adults—scientists, musicians, business entrepreneurs, politicians, craftspeople, attorneys, novelists, physicians, designers, and so on. What are the characteristics of their work that justify calling their accomplishments authentic rather than contrived and trivial? Can we identify key distinctions between these authentic accomplishments and the work that students complete in school?[2]

Authentic achievement involves the challenge of producing, rather than reproducing, knowledge.

People in the diverse fields named above face the primary challenge of producing, rather than reproducing, knowledge. This knowledge is expressed through discourse, through the creation of things, and through performances. We do not expect children to attain levels of competence comparable to those of skilled adults, but we do want students to develop in the same direction.

To progress on the journey toward adult competence, students should set their sights on authentic expressions of

knowledge. That is, they must hone their skills through guided practice in producing original conversation and writing, through the repair and building of physical objects, through artistic and musical performance. The conventional curriculum asks students only to identify the discourse, things, and performances that others have produced—e.g., by recognizing the difference between verbs and nouns or between socialism and capitalism; by matching authors with their works; or by correctly labeling rocks and body parts. Certainly, the production of knowledge must be based on an understanding of prior knowledge, but the mere *re*production of that knowledge does not constitute authentic academic achievement.

A second defining feature of authentic academic achievement is its reliance on a particular type of cognitive work that can be described as "disciplined inquiry." Disciplined inquiry, in turn, seems to consist of three features: 1) use of a prior knowledge base, 2) in-depth understanding rather than superficial awareness, and 3) production of knowledge in an integrated (rather than fragmented) form. In highlighting these features, I am not suggesting that students should be expected to make seminal contributions to the academic disciplines, to the professions, or to the arts. However, I am suggesting that students are quite capable of engaging in these forms of cognitive work when they are adapted to the students' levels of development.

1. **Prior knowledge base.** For new knowledge to be significant and valid, it must be based on substantive and procedural knowledge that has been accumulated previously by workers in the field, who establish facts, vocabularies, concepts, theories, algorithms, and conventions for the conduct and expression of inquiry. However, relying on a prior knowledge base does not always mean agreeing with intellectual traditions or continuing to develop them in a linear fashion. The ultimate point of disciplined inquiry is to move beyond existing knowledge—through criticism and the development of new models that are themselves stimulated by prior knowledge. Most of the cognitive work of schools consists of transmitting prior knowledge to students and asking them to reproduce it,

rather than helping them to use it to produce new knowledge.

2. **In-depth understanding.** Disciplined inquiry tries to develop in-depth understanding of a problem, rather than only a passing familiarity with or exposure to pieces of knowledge. One masters prior knowledge not primarily to become superficially knowledgeable about a broad range of topics, but in order to develop complex understanding of a smaller number of issues. Such detailed understanding is particularly necessary for the production of new knowledge. In contrast, many of the cognitive tasks of school ask students to show only superficial awareness of a vast number of topics.[3]

3. **Integration.** To produce knowledge, one must assemble and interpret information, formulate ideas, and make critiques that cannot be easily retrieved from the existing knowledge base. All these activities require the ability to organize, synthesize, and integrate information in new ways. Students are unlikely to succeed at such tasks unless they learn to look for, to test, and to create relationships among pieces of knowledge that otherwise appear unconnected.

The third and final distinction between authentic achievements and traditional school achievements is that authentic achievements have aesthetic, utilitarian, or personal value apart from their value in documenting the competence of the learner. Indeed, this distinction may be the most crucial. When people write letters, news articles, insurance claims, or poems; when they speak a foreign language; when they develop blueprints; when they create a painting, compose a piece of music, or build a stereo cabinet, they try to communicate ideas, to produce a product, or to make an impression on others beyond the simple demonstration that they are competent. Achievements of this sort have special value that is missing from tasks contrived only for the purpose of assessing knowledge, such as spelling quizzes, laboratory exercises, or typical final exams. In many cases, the cry for "relevance" is simply a less precise expression of the desire

Accomplishments should have intrinsic value beyond their value as indicators of success in school.

that accomplishments should have intrinsic value beyond their value as indicators of success in school.

To summarize, the idea of authentic achievement requires students to engage in disciplined inquiry to produce knowledge that has value in their lives beyond simply proving their competence in school. Mastery of this sort is unlikely to be demonstrated in familiar testing and grading exercises. Instead, such mastery is more often expressed in the completion of long-term projects that result in the creation of discourse, things, and performances of interest to students, their peers, and the public at large.

Why should we aim toward authentic achievement? I suggest at least two compelling reasons, which respond directly to the problems of the conventional curriculum that I mentioned earlier. First, participation in authentic tasks is more likely to motivate students and to sustain the hard work that learning requires. Because authentic work has value beyond the mere demonstration of competence in school and because it permits more comprehensive use of the mind, students will have a greater stake in authentic achievement. Second, authentic academic challenges are more likely to cultivate the higher-order thinking and problem-solving capacities that are useful both to individuals and to society. The mastery gained in school would then transfer more readily to life beyond school and thus increase the efficiency of our public investment in schooling.

To develop organizational structures consistent with this vision, educators will need to plan "backwards." Beginning with a conception of authentic achievement and of the specific projects that express it, they will need to select appropriate content within the subjects. They will also need to consider the kinds of human interaction—involving teachers, students, and others—that promote disciplined inquiry into the subject. Although I cannot address organizational needs for specific subjects in this article, one aspect of interaction seems critical to authentic achievement in many subjects. This is the need for substantive conversation in instruction, which constitutes a major departure from present practice.

SUBSTANTIVE CONVERSATION

When the curriculum is geared toward the goal of authentic achievement, new forms of teaching will be necessary. No longer will it be possible for students to succeed in school simply by listening to the teacher and responding to questions that can be answered in half a dozen words. To complete authentic projects, students will need to formulate and to ask questions, to explain themselves to peers and to adults, and to refine their ideas. While many authentic achievements require capacities other than verbal discourse (e.g., working with one's hands, making mathematical estimations, using aesthetic sensitivities), substantive conversation between the learner and some supportive resource—teacher, peer critic, or other knowledgeable authority—is usually necessary. I emphasize substantive conversation, not because it alone is sufficient to develop all forms of authentic achievement, but because it is usually critical for success in authentic tasks and because conversation is the primary medium through which teachers and students communicate.

When the curriculum is geared toward the goal of authentic achievement, new forms of teaching will be necessary.

Substantive conversation differs from conventional classroom talk, in which the teacher usually has only two purposes (aside from disciplinary ones) in speaking to students: to transmit an item of declarative knowledge (usually a definition or a fact) or to determine whether the student can reproduce that item.[4] The student's goal (assuming that he or she desires to cooperate) is to give the answer that the teacher has in mind. The teacher maintains control in that he or she knows, prior to the interaction, what pieces of knowledge must be communicated and what answers will be acceptable.

In a substantive conversation, the purpose of talking is quite different. Each person is trying to express a point of view (often in order to persuade the other), to understand why the other holds a particular viewpoint or interpretation, or to arrive at a solution to a problem that neither has previously solved. This last purpose is the most difficult to fulfill, but it is perhaps the most likely to render the work authentic.

In pursuing any of these ends, the teacher uses talk to integrate subject matter into the student's meaning system, and the student tries to understand how the teacher's messages might empower him or her in the world. In this sense, each is interested in what the other has to say, the course of the conversation is less predictable than in conventional classroom exchanges, and each participant has some sense of ownership. To fulfill these purposes, substantive conversation requires sustained, continuous talk between two or more people. In school as we know it, the two key participants are usually student and teacher, but students can also learn through conversations with their peers, with other adults, and perhaps with well-programmed computers. Teacher/student conversation should remain central. But to expand students' opportunities for expression and feedback, it will be necessary to rely substantially on and to support these other sources as well.

What is the connection between substantive conversation and authentic achievement? Begin with the definition: authentic achievement is expressed in discourse, things, and performances that make reasonably complete, integrated statements and that reflect students' production of in-depth knowledge. Then ask the question, How can we instruct students to make such statements? We can show them models of authentic accomplishments of others, and we can give them critical elements of declarative knowledge on which to build. But how will they learn to make their own statements? Substantive conversation is the key. It provides the crucible for practice, for seeking new knowledge that relates to the problem at hand, and for trial, feedback, and revision. In short, substantive conversation forces us to transform declarative knowledge into applied, integrated knowledge.

STRUCTURES THAT FACILITATE AUTHENTIC ACHIEVEMENT

The Council of Chief State School Officers divides restructuring into four main categories: school governance (including decentralized authority and school choice), reforming the nature and organization of curriculum and instruction, new professional roles for educators, and

accountability (especially new methods of assessment and direct state intervention).[5]

This analysis applies most directly to curriculum, instruction, and accountability. But I believe that, unless the entire restructuring agenda focuses more closely on authentic student achievement and on substantive conversation to attain it, restructuring will produce at best only cosmetic changes in students' education.

Assuming a commitment to these goals, what structural conditions seem necessary to achieve them? The conditions I propose below are in no way exhaustive. They outline certain key conditions under which students would work, but they do not provide structures of school governance, detailed systems of assessment, or specific new structures for the professional development of teachers. My argument has implications for these areas as well, and I do support the need for providing teachers with working conditions that enable new forms of professional support. But my central concern here is the structure of the instructional environment.

Useful collaboration depends to a great extent on opportunities for substantive conversation, which itself requires special structural conditions.

At least four conditions seem essential for students to complete authentic accomplishments: collaboration, access to tools and resources, worker discretion or opportunity for ownership, and flexible use of time.

Collaboration. Achievements outside of school often depend on the opportunity to ask questions of, to receive feedback from, and to count on the help of others, including peers and authorities. In contrast, typical school activities require each student to work alone. Working together is often prohibited, because it is seen as a form of cheating. Of course, it is important for students to learn to work on their own and not become too dependent on others. But if opportunities to cooperate are consistently denied, students will not experience a critical process that adults, both expert and novice, consistently rely on for success.

Useful collaboration depends to a great extent on opportunities for substantive conversation, which itself requires special structural conditions. More time will be

needed for teachers to communicate with individual students through sustained talk and writing and for students to talk with one another. Substantive conversation also entails major shifts in the roles of teachers and students. Teachers will function more as mentors and coaches, less as depositories of static knowledge to be reproduced. Students will function more as constructors and producers of knowledge. They will rely on teachers for help, but they will not be mere absorbers or consumers of everything the teacher says. Students will also have to take on the new roles of seeking help from and giving help to one another as they learn.

Access to tools and resources. When competent people complete significant accomplishments—e.g., writing books, designing buildings, composing music, or reaching collective bargaining agreements—they usually have made use of a variety of printed materials and such machines as telephones, computers, and laboratory equipment. The information they require is generally too vast and too rapidly updated to commit to memory. In medieval times, before the invention of the printing press, becoming educated required memorizing the contents of key manuscripts, because future access to them was unlikely. But today increasing specialization and the explosion of knowledge create a new challenge. Possession of prior knowledge is critical to disciplined inquiry, but it is now impossible to teach students all the relevant information they will need to complete authentic tasks. The new challenge is to teach students to use tools and resources to find the appropriate knowledge when they need it.

Discretion and ownership. Rather than always toiling within the confines of predetermined routines arbitrarily dictated by authorities, the creators of authentic work influence the conception, execution, and evaluation of the work itself. At a minimum this entails exercising some control over the pace and procedures of learning; over opportunities to ask questions and to study topics deemed important; and over constructing and producing knowledge in one's own language, rather than merely reproducing the language of others.

There are, of course, important limits on the extent to which students should control their learning of academic subjects. Certain facts, definitions, concepts, algorithms, and processes of verification must be assimilated according to the standards of the disciplinary fields. But for this kind of learning to be translated into authentic products, students must enjoy some autonomy and discretion in its application.

Flexible use of time. The significant achievements of disciplined inquiry often cannot be produced within rigidly specified time periods. Adults working to solve complicated problems, to compose effective discourse, or to design products are rarely forced to work within the rigid time constraints imposed on students, such as the 50-minute class or the two-hour examination.

Standard time schedules stem from bureaucratic procedures designed to manage masses of students and diverse course offerings.

Standard time schedules stem from bureaucratic procedures designed to manage masses of students and diverse course offerings. Since they have nothing to do with the time requirements of disciplined inquiry, they can reduce the authenticity of student achievement. Achievements in noninstructional tasks, such as journalistic writing, interior design, or medical care, do involve deadlines and time limits. But in such work the schedules tend to be determined more by the nature of the work than by the requirements of institutional management.

The four conditions outlined above suggest at least the following changes in the structuring of students' work.

- To stimulate collaboration and the substantive conversation so crucial to it, more time must be allocated to individualized and small-group instruction.
- Students' access to knowledge must be enhanced by greater use of technology (telephones as well as computers) and by opportunities to learn from sources outside of school.
- Students need to exercise discretion and assume responsibility in the planning, execution, and evaluation of their work, which further underscores

the need for individualized and small-group work along with assessment that accepts diverse ways of demonstrating competence.

- Instructional time should be organized to permit more sustained, long-term, and in-depth investigation, in contrast to the fixed time slots designed for survey coverage.
- Assessment tasks calling for the production of discourse, things, and performances will require new procedures and in some cases the participation of individuals who are not members of the school staff to evaluate student performance.

Instituting these changes will contribute to authentic achievement. But they alone cannot guarantee it, for ultimately the quality of student achievement will depend on the quality of conversation that students have with teachers and with one another. For years substantive conversation has been difficult to find in schools, but only partly because of structural constraints. Even if teachers were blessed with more time and much lighter teaching loads, I suspect that substantive conversation would still be rare, because the goals of authentic achievement and substantive conversation also presume major shifts in prevailing conceptions of education. Organizational changes alone will not modify long-standing conceptions of knowledge and deeply socialized habits of didactic teaching that direct most of our educational efforts toward inauthentic forms of mastery.

We must resist the tendency to institute administrative changes devoid of fresh educational vision.

Just as plans for restructuring should begin with the conditions in which students learn, the next step is to consider the working conditions that teachers need in order to generate substantive conversation. Again, we must not expect organizational structures alone to transform teaching. However, given the complexities of the teaching task, we cannot emphasize enough the need to provide teachers and administrators with sufficient authority, time, and assistance to reflect critically on their practice, to experiment, to fail, and to try again. Beyond enabling teachers to increase the proportion of time they spend interacting with individual students and small

groups, we must work to give teachers more opportunity to collaborate, and we must provide for regular, sustained professional development that focuses on the goals of substantive conversation and authentic achievement.

Experiences in restructuring to date reveal three obstacles that dramatize the need for this kind of support. First, teachers, parents, and students who have experienced only the conventional version of education—especially those who have been successful—cling tenaciously to it, even when they have the opportunity to make substantial changes. Faith in conventional practice is fortified by the large proportion of students who do succeed on conventional measures, both in school and beyond. Second, external pressures for district, state, and national accountability, coupled with the need for ways to determine admission to college, continue to sanctify mastery of declarative knowledge as the most important goal of schooling. Finally, the struggle to establish new collaborative roles for teachers is stressful and time consuming; it drains energy and leaves diminished personal resources for thinking about new forms of achievement and curriculum.

The history of education reform suggests that we have yet to see fundamental change on a grand scale.[6] We might argue that the contemporary sense of urgency and the scope of the current restructuring movement combine to offer a unique opportunity. But to avoid the costly mistakes of previous efforts, we must resist the tendency to institute administrative changes devoid of fresh educational vision. If we keep our sights on the targets of authentic student achievement and substantive instructional conversation, restructuring could form a new chapter in educational history—and one well worth the effort.

This article is adapted from a speech presented at the Indiana University Annual Education Conference, held in Bloomington in April 1990, and its writing was supported by Indiana University, by the Office of Educational Research and Improvement of the U.S. Department of Education, and by the Wisconsin Center for Education Research. However, the opinions expressed are the author's own.

NOTES

1. Albert Shanker, "The End of the Traditional Model of Schooling—and a Proposal for Using Incentives to Restructure Our Public Schools," *Phi Delta Kappan*, January 1990, pp. 344-57.
2. This complicated matter involving controversial educational values has not yet received extensive scholarly attention. Yet it begs for interdisciplinary analysis of the relationship between formal attempts to educate and the ways that human competence is expressed in noneducational settings. The conception I propose here builds on Harold Berlak, ed., *Toward a New Science of Educational Testing and Assessment* (Albany: State University of New York Press, forthcoming); Fred M. Newmann and Doug Archbald, *Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School* (Reston, Va.: National Association of Secondary School Principals, 1988); John Raven, "Toward a Model of Competence, Motivation, and Behavior and a Paradigm for Assessment," In Berlak, op. cit.; Lauren B. Resnick, "Learning in School and Out," *Educational Researcher*, December 1987, pp. 13-20; and Grant Wiggins, "A True Test: Toward More Authentic and Equitable Assessment," *Phi Delta Kappan*, May 1989, pp. 709-13.
3. A commitment to depth over coverage entails no retreat from liberal education into narrow forms of vocationalism, as some might expect. The effort to concentrate on complex understanding welcomes study of a wide variety of disciplines and paths to knowledge. Reducing the emphasis on superficial exposure implies no necessary narrowing of the curriculum as a whole.
4. This analysis is based in part on the work of Martin Nystrand and Adam Gamoran, "Student Engagement: When Recitation Becomes Conversation," in Hersholt Waxman and Herbert Walberg, eds., *Contemporary Research on Teaching* (Berkeley, Calif.: McCutchan, forthcoming).
5. *Success for All in a New Century: A Report by the Council of Chief State School Officers on Restructuring Education* (Washington, D.C.: Council of Chief State School Officers, 1989).
6. Larry Cuban, "Reforming Education, Again, Again, and Again," *Educational Researcher*, January/February 1990, pp. 3-13.

BEST COPY AVAILABLE

What Teachers Learn from "Kid Watching"

by Richard VanDeWeghe

Teacher-researchers in a Colorado district are expanding their understanding of teaching and learning by viewing their classrooms as "thoughtful communities," where they, too, are learners.

Recently I asked teachers participating in the Douglas County School District's Higher Literacy Project about their classroom-based research. One teacher talked excitedly about "an absolutely amazing event that happened just this morning." She then shared an anecdote about an at-risk 8th grader who wrote something longer than a sentence for the first time since September: a poem about his alienation from school. For this teacher-researcher, an unexpected event, like the poem suddenly and awkwardly composed by a marginally literate child, made her ask questions: Why did he do that now? What let loose in him? Has he learned to trust me? The class?

What makes this project distinctive is the individually designed programs of professional growth determined by the teacher-researchers themselves. To commit to doing research constitutes a major investment in change.

Another teacher looked down at his feet, hands in pockets, and admitted that his project is "taking turns I didn't expect" and that "the more I listen—really listen—to how these kids talk about their reading in the history textbook, the more I think I need to know about *reading*!" For him, realizing that he knows little about the problems kids have with reading history texts is seen as a professional opportunity rather than a shortcoming.

From *Educational Leadership*, vol. 49, no. 7, p. 49-52, 1992. Reprinted with permission.

Another response came from a teacher who suddenly reached into her skirt pocket, pulled out a small notebook, flipped over pages of penciled notes, and read her observations of a class debate about a Hemingway short story. For her, getting some insight into the patterns of discourse in class debates is exciting.

These are typical responses of teacher-researchers who view their classrooms as places where they are the learners—learning from their students. As Mary W. Olson said recently, "These are confident teachers who research questions that intrigue and puzzle them, who seek answers and understanding about their students' learning and their own teaching, and who strive to be more knowledgeable in their responses to the teaching/learning cycle" (1990, p. 13).

THOUGHTFUL COMMUNITIES

Teacher research in Douglas County, Colorado, is sponsored by the district's Thoughtful Communities: The Douglas County Higher Literacy Project. Participants commit to conducting two research projects in two years. They form "thoughtful communities" with other district teacher-researchers, who collaborate with and support one another in their research. At the same time, they seek to create thoughtful communities of learners in their classrooms: groups of students who collaborate with one another in constructing and refining their own knowledge. The second part of the program's title indicates its aim to promote and understand the nature of higher literacy, that complex of skills, attitudes, and abilities that mark the genuinely literate person of the 21st century (Brown 1988).

What makes this project distinctive is the individually designed programs of professional growth determined by the teacher-researchers themselves. To commit to doing research (in addition to all the other things that draw on their time and energy) constitutes a major investment in change. Then there is the large-scale district support for teachers-as-researchers: participating teachers receive release time for project research, university graduate credit, hourly pay for attending meetings, books to add to their professional libraries, and expenses paid for project-related professional conferences.

Project teachers meet regularly to discuss readings, issues and problems in higher literacy and classroom research; help one another (in writing workshops) to prepare project reports; and participate in institutes related to their professional development.

THE RESEARCH PROJECTS

By observing their students at work, asking good questions about learning, and, as Glenda Bissex says, seeing "everything that happens in a classroom...as data to be understood rather than causes for blaming or congratulating" (1986, p. 483), Thoughtful Communities teachers take on research projects that illuminate their understanding of teaching and learning, regardless of the projects' potential for success or failure. "We can learn from failures as well as successes, same as our kids," one teacher recently remarked. A sampling of their projects includes:

- investigation of ways to help students learn to write in visually articulate ways (using graphic, typographic, and spatial techniques);
- inquiry into the contradictions teachers face when they take time to promote higher literacy in their classrooms;
- development of a writing portfolio approach to outcomes-based assessment in social studies;
- exploration of the learning value of ambiguity and speculative thinking in language arts and social studies;
- development of classroom structures that address the needs of at-risk students.

The projects originate from a number of sources, and in each source there exists some kind of tension. One source is *pedagogical experimentation*, where the tension is inherent in such speculation as "I wonder if this approach could work for me?" This was the case with one of our researchers, who came into the project intrigued by a teaching approach that encouraged active questions. He determined that he wanted to find out if, how, and why active questions contribute to higher literacy.

The second source of tension has to do with *self-reflection* and the corresponding doubt that gives rise to

action. Like other reflective educators, project teachers engage in a "cycle of thought and action based on professional experience" (Wellington 1991, p. 4). For example, many teachers change completely their initial project area once they begin reading and talking with others about higher literacy. They say, in effect, "Why am I not promoting higher literacy in the way I teach?" The answers to that single question lead them to reflect on how they teach, how their kids learn, and how things could be.

A wonderful sense of empowerment accrues to teachers who have a database for their teaching, who can say, "I know this works; here, let me show you these few hundred writing samples to illustrate what I mean."

Third, teachers *discover* projects through their research activities. They may tentatively commit to a project and in the course of reading about their topic, collecting classroom data, and reflecting on both, find their *real* project topic. They say, "Knowing *this* leads me to think in these ways now." This is what happened with a language arts teacher whose research topic began as an investigation of nonstandardized methods of assessing reading and writing. After seven months, her focus shifted to developing a learning environment in her classroom where assessment became integrated with, rather than separated from, her teaching.

Much professional growth occurs through trying to understand or resolve these tensions. Teachers discover that new techniques do or don't work for them—and because they are experimenting thoughtfully, they get insight into why or why not. That's much different from the teacher who tries and fails and never asks why. Similarly, teacher-researchers who doubt themselves or their approaches rely on their research to prove or disprove the legitimacy of their doubt. And those who think new thoughts reconceptualize their research because they know that doing so is a primary reason for such research to go on.

Classroom-research-as-professional-development becomes the impetus as well as the vehicle for change. I saw this idea brought out in one teacher's research report:

> As an observer, I was able to examine "why" and "how" things happen as they do. For the first time ever I brought

> questions as well as answers to my classroom; questions that have helped me better evaluate not only my students but also myself. In a relatively short time I have made, and plan to continue making, significant changes in my classroom. My sense is that without a component of teacher-based research, change would be slow in coming (Groll 1990, p. 37).

TEACHERS REFLECT ON THEIR GROWTH

The best indicators of professional development in this project come from the teacher-researchers themselves, as they've contemplated their experiences in Thoughtful Communities. Here are just a few examples.

One teacher described how conducting classroom research altered the way she viewed herself:

> I experienced many successes and frustrations along with some startling perceptions of myself as a teacher.... My successes include being a part of a happier, more exciting classroom experience with a minimum of "policewoman" responsibilities and the realization that it's okay not to be the ultimate expert in the eyes of my students.

Another teacher, whose project focused on the art of questioning, ended up with a rich and vastly expanded understanding of the heart of classroom inquiry:

> Questions are not only a way of gaining knowledge—questions and knowledge fission to generate more questions and more knowledge. Questioning isn't a method or a technique; it has to be a habit—a lifestyle—if it is to lead students to higher levels of literacy.

Finally, a teacher whose research changed his view of school writing from "writing to present ideas" to "writing to learn" found that in addition to promoting a view of writing-as-a-way-of-learning, he needed to promote talking-as-a-way-of-learning:

> I need to develop a skill in questioning that allows students to think. In the past, I usually orchestrated the discussion to get the answers I wanted to hear.

INSIGHT INTO PROFESSIONAL DEVELOPMENT

What does teacher research of the kind described here teach us about professional development? First, *true* professional growth accompanies a skeptical stance toward oneself and toward one's classroom; it's a form of critical curiosity that many nonresearchers find hard to fathom.

Second, a collaborative network of teachers engaged in similar growth nourish and sustain one another in vital ways, especially during times of great challenge or doubt.

Third, collecting data—for example, observing and listening to kids, taking notes, reading learning logs and writing samples—forces teachers into the practice of monitoring learning in ways that demand reflection. That is, if things aren't going as one expects, one must ask why. Also, a wonderful sense of empowerment accrues to teachers who have a database for their teaching, who can say (with great gusto), "I know this works; here, let me show you these few hundred writing samples to illustrate what I mean."

Finally, teacher-researchers challenge systems (curriculum, assessment, for example). But they do so from affective and cognitive knowledge based on firsthand observation, wide reading, and deep conviction.

Teacher-researchers constantly probe and question, listen and observe, notice and note—to them, professional development is driven by the vision, knowledge, and ambition generated by their research. They are, in Yetta Goodman's words, "kid watchers" (1989, p. 8), teachers who interact with students and who monitor class activities in order to understand more about teaching and learning, mostly learning.

REFERENCES

Bissex, G. (1986). "On Becoming Teacher Experts: What's a Teacher-Researcher?" *Language Arts* 63, 5: 482-485.

Brown, R. (Spring 1988). "Schooling and Thoughtfulness." *Basic Education* 3, 3.

Goodman, Y. M. (1989). "Evaluation of Students." In *The Whole Language Evaluation Book*, edited by K. S. Goodman, Y. M. Goodman, and W. J. Hood. Portsmouth: Heinemann.

Groll, L. (1990). "Thoughtful Communities: A Reflection." In *Thoughtful Communities: Douglas County Higher Literacy Project.* Castle Rock, Colo.: Douglas County School District Reorganized 1.

Olson, M. (June/July 1990). "Teachers as Researchers." *Reading Today* 7, 6: 13.

Wellington, B. (1991). "The Promise of Reflective Practice." *Educational Leadership* 48, 6: 4-5.

Assessment: A Thoughtful Process

by Steven Ferrara and Jay McTighe

Many teachers are seeking to involve their students in the meaningful application of knowledge and skills. These teachers seek to create a thoughtful classroom in which students assume a greater responsibility for constructing meaning for themselves.

The development of student thinking abilities has been recognized as a basic educational goal for the information age (Secretary's Commission on Achieving Necessary Skills, 1991; McTighe & Schollenberger, 1991). With this recognition in mind, many teachers are seeking to involve their students in the meaningful application of knowledge and skills, rather than simply having them memorize facts. These teachers seek to create a thoughtful classroom in which students assume a greater responsibility for constructing meaning for themselves. Such a classroom emphasizes active learning, involving students in working together to examine issues, solve problems, and communicate ideas. The teacher adopts a more facilitative role, guiding students in individual and group work, rather than merely dispensing information.

The classroom context offers a distinct advantage over large-scale assessments in that it allows teachers to take frequent samplings of student learning in different contexts.

In addition to changing classroom expectations and the role of teachers and students, such an instructional orientation has implications for the ways in which student learning is assessed. A greater emphasis is placed on the use of assessments that involve students in more "authentic"

From *If Minds Matter: A Foreword to the Future, Vol. II*, Arthur L. Costa, James Bellanca, and Robin Fogarty, eds., 1992, p. 337-348. Reprined with permission.

tasks[1], measure a variety of important learning outcomes, and reflect effective instructional models. Unfortunately, many teachers find that they have received little relevant training in methods of classroom assessment (Stiggins, 1991), either through university coursework or inservice workshops.

In this chapter, we focus on the classroom as we examine principles of sound classroom assessment, introduce a planning process and a framework for selecting appropriate assessment methods, and provide illustrations of the process and framework in use. Rather than examining specific assessment methods in detail, our concentration is on helping teachers develop thoughtful classroom assessments.[2]

PRINCIPLES OF SOUND CLASSROOM ASSESSMENT

The term *assessment* refers to "any systematic basis for making inferences about characteristics of people, usually based on various sources of evidence; the global process of synthesizing information about individuals in order to understand and describe them better" (Brown, 1983, p. 485).[3] Inherent in this definition is an important principle, namely that sound classroom assessment practice requires multiple sources and types of data. Applied to the classroom assessment context, this principle reminds us to rely on more than paper-and-pencil tests to gauge student learning.

To examine the principle of multiple data sources more fully, let's use a photographic analogy. Think of an experience in which someone took a candid photograph of you. You may have been happy if it "captured" you in a pleasing manner. On the other hand, perhaps you were a bit embarrassed because the image was not particularly flattering. In fact, it may have been downright awful, and you may have attempted to get it back so that others would not see it! In either case, would you be willing to say that this single photograph effectively portrays you? At best, it presents a picture of you at a single moment in time within one particular context. In a similar way, the result of a single assessment, be it from a large-scale assessment (district, state, or national test) or an end-of-unit classroom test, provides incomplete information. It

would be inappropriate to use such a one-time snapshot of student performance as the sole basis for drawing conclusions about how well a student has achieved desired learning outcomes.

The classroom context offers a distinct advantage over large-scale assessments in that it allows teachers to take frequent samplings of student learning in different contexts. To continue the photographic analogy, classroom assessment enables us to construct a "photo album" containing a variety of pictures taken at different times with different lenses, backgrounds, and compositions. The photo album reveals a richer and more complete picture of each student than any single snapshot provides. Similarly, effective classroom assessment requires multiple sources of data, gathered over time, regarding student learning to enable us to make sound inferences about what our students know and can do.

What teachers assess and how they assess should be congruent with their desired learning outcomes.

The use of multiple sources of assessment data is important for summative assessments that form a basis for assigning report card grades. Rather than relying on a single end-of-unit test for assigning grades, effective teachers factor together a variety of sources of information, such as student products, performances, class participation, homework, and conferences. In addition, teachers can use multiple sources of data as the basis for formative assessments. For example, prior to the start of a unit on the Civil War, the teacher might ask students to make a web showing what they already know about this period of history as a means of obtaining information about students' prior knowledge. The teacher might also interview several students at random to check their Civil War awareness. Formative assessment can also be used during instruction to check on student understandings and misconceptions. For instance, teachers often use brief written and oral quizzes and classroom discussions to determine if students have learned course material and can apply instructed skills. Such activities provide teachers

with invaluable feedback on their teaching and allows them to adjust instruction to achieve more effective learning.

A second principle of sound classroom assessment practice relates to the relationship between instruction and assessment. It asserts that what teachers assess and how they assess should be congruent with their desired learning outcomes and the instructional methods they use to reach these outcomes. Learning outcomes specify what teachers want their students to know and to be able to do as a result of their educational experiences (Spady and Marshall, 1991). Learning outcomes may consist of knowledge (e.g., the Pythagorean theorem), skills (e.g., a problem-solving procedure), attitudes (e.g., appreciation of literature), or "habits of mind" (e.g., concern for accuracy). Instructional methods may include teacher presentations, cooperative learning, hands-on problem-solving activities, and independent study.

According to this "congruence principle," classroom assessments should be aligned with outcomes and instruction so as to provide valid measures of learning. For example, if teachers want students to demonstrate the capacity to write an effective persuasive essay, then their assessment would include samples of such writing evaluated against specified criteria. The use of a multiple-choice exam, in this case, would be an ill-suited measure of the intended outcome. Likewise, if teachers wish to develop students' ability to work collaboratively on a research project, then they would want to assess group processes and products as well as individual performance. Aligning assessment with curriculum and instruction is crucial because what and how teachers assess sends a clear message to students about what knowledge, skills, processes, and attitudes are most important for them to learn and develop.

SELECTING CLASSROOM ASSESSMENT METHODS

As with instruction, a variety of assessment approaches are available. The selection of methods for gathering information on students' learning should be determined, in part, by the congruence of these methods with desired out-

comes and instructional approaches. In addition, teachers should consider at least three other factors when selecting assessment methods. First, assessment methods should be appropriate to the purpose(s) of, and the audience(s) for, a classroom assessment. For example, norm-referenced tests, such as the Iowa Tests of Basic Skills, are used primarily to satisfy demands for educational accountability. Their results are reported to the legislature, the Board of Education, school administrators, parents, and the general public. They are generally designed to determine how well students have learned particular concepts and skills as compared to other students in a norm group. However, at the classroom level, teachers are generally more interested in using assessments for diagnosing student needs, informing students and parents about progress, adjusting instruction, and motivating students. These varying purposes and audiences should determine the methods of assessment and the means by which the results are communicated.

At the classroom level, teachers are generally more interested in using assessments for diagnosing student needs, informing students and parents about progress, adjusting instruction, and motivating students.

A second consideration concerns the issue of reliability. Reliability refers to the dependability and consistency of assessment results. For instance, a writing assessment would be considered reliable when different raters assign similar scores to the same essays. Likewise, an observation checklist can be used reliably as long as teachers are careful to recognize that their ratings may differ from occasion to occasion (e.g., Monday morning versus Friday afternoon), and that their perceptions of individual students may be influenced by their attitudes toward each student.

Finally, practicality must be considered when selecting assessment methods for the classroom. For example, while a science fair project/exhibition may provide a valid and reliable means of assessing student learning in a science class, the time required to prepare and present the project and the space needed for displaying the exhibit may be difficult to obtain in some schools.

A FRAMEWORK OF ASSESSMENT APPROACHES				
SELECTED RESPONSE Formats	**CONSTRUCTED RESPONSE** Formats	**PRODUCT** Assessments	**PERFORMANCE** Assessments	**PROCESS-FOCUSED** Assessments
1. multiple choice 2. true/false 3. matching	4. fill in the blank 5. label a diagram 6. short answer 7. "show your work" 8. concept map 9. figural representation	WRITTEN 10. essay, story, or poem 11. research report 12. writing portfolio 13. diary or journal OTHER 14. science fair project 15. art exhibit or portfolio	16. musical, dance, or dramatic performance 17. science lab demonstration 18. typing test 19. athletic competition 20. debate 21. oral presentation	22. oral questioning 23. interview 24. think aloud 25. learning log 26. process folio 27. kid watching

Figure 1: A Framework of Assessment Approaches

THE FRAMEWORK OF ASSESSMENT APPROACHES

How then do we go about selecting appropriate assessment approaches to use in our classrooms? The Framework of Assessment Approaches[4] (Figure 1) offers a systematic guide to help select and develop assessment methods for the classroom.

Assessment Approaches

Each of the five columns in the framework presents an assessment approach or format. Organizing the five approaches in this way highlights the most important similarities of the methods in the same column, and the most useful distinctions between the methods in different columns. However, the methods could be categorized differently. For example, "essays" could be listed under "constructed response formats," and "oral questioning" might be included under "performance assessments" or "constructed response formats."

The most widely used approaches to classroom assessment appear in the first and second columns. In **selected response formats**, students select the one correct or best answer from several alternatives, for example, multiple-choice items on the Comprehensive Tests of Basic Skills. In **constructed response formats** students generate a response (usually short) to an open-ended question, problem, or task, for example, when students are asked to label a diagram or chart, explain their reasoning, or show their work. However, growing recognition of the limitations of these approaches has led to calls for richer, more "authentic" assessment approaches (National Commission on Testing and Public Policy, 1990), like those listed under the third, fourth, and fifth columns. Teachers may use **product assessments** in which students create tangible products, such as a research paper, science fair project, or art portfolio. They may also use **performance assessments** in which students engage in actual performances that require demonstrations of desired knowledge and skills, for example, oral presentations, demonstrations, or use of an online database to gather information. We draw a distinction between product and performance assessments to highlight the difference between evaluating a performance itself (e.g., students conducting a science lab demonstration) and evaluating the product of a performance (e.g., a science lab report). We further distinguish **process-focused assessments**, in which teachers elicit students' thinking processes (e.g., process observing, "think alouds"), because they are explicitly intended to provide insight into students' thinking processes rather than the products and performances that result from their thinking.

Assessment Methods

The methods listed in each column of Figure 1 are intended to characterize each assessment approach rather than exhaustively define it. Some of these methods are clearly associated with testing, such as the multiple-choice items. Less formal methods, such as oral questioning and "kid watching," may not necessarily be associated with classroom assessment (Haertel, Ferrara, Korpi, & Prescott,

1984). However, use of these informal methods is essential to providing a complete picture of student learning. It may be useful to provide examples of these methods. Methods one through five in the framework are generally familiar to teachers because they are among the most widely used classroom assessment methods. They are probably most often used to find out whether or not students can recall and understand factual information.[5]

Determining what knowledge and thinking skills and processes students apply in classroom learning situations requires classroom assessment methods that focus directly on thinking processes.

In contrast, teachers use other assessment methods to get beyond the "correct" answer to more open-ended tasks in which an array of responses is acceptable and expected. Teachers often include short-answer questions in their tests (method six), for example, to pose questions that may not have a single correct or best answer. Math teachers may require students to "show their work" (method seven) to determine what steps students follow in solving multi-step problems. Concept maps, semantic webs (Novak and Gowin, 1984), and other figural representations (methods eight and nine) allow teachers to gain insight into how students organize facts and concepts and to gauge the accuracy and complexity of their understandings.

Teachers assign essays and longer writing assignments such as research reports (methods ten and eleven) to provide feedback about students' abilities to locate and organize information and express ideas clearly in written form. In recent years, we have witnessed increased use of portfolios (method twelve) in language arts and other subject areas to collect representative student work and to chronicle development over time. Product exhibitions, such as science fair projects and art exhibits (methods fourteen and fifteen), allow students to demonstrate their ability to apply knowledge and skills in authentic contexts.

Performance assessments are especially well suited to certain content areas since they allow teachers and others to directly observe the application of desired skills. For example, teachers grade or critique students as they

perform a musical piece, demonstrate a laboratory procedure, type for speed and accuracy, or participate in an athletic competition (methods sixteen through nineteen). Teachers may also organize debates and assign oral reports (methods twenty and twenty-one) in order to evaluate students' performances.

Determining what knowledge, thinking skills, and processes students apply in classroom learning situations requires classroom assessment methods that focus directly on thinking processes. Teachers may elicit students' thinking processes using oral questions like "How are these two things alike and different?" (method twenty-two); by interviewing students (method twenty-three) regarding the procedures they follow to generate products or what they think about while performing a task; or by asking students to "think out loud" (method twenty-four) as they solve a problem or make a decision. In addition, teachers may ask students to document their thinking processes by keeping a diary, learning log (method twenty-five), or process folio (method twenty-six). Finally, teachers can learn about students' thinking processes by watching students individually and in groups as they function in the classroom—that is, by "kid watching" (method twenty-seven). This method is especially well suited to assessing the development of habits of mind, such as persistence and open-mindedness.

We provide these brief examples to highlight the variety of classroom assessment methods available to teachers. How then do we decide which approaches and methods to use? How can we make classroom assessments more thoughtful, sound, and effective? The following process offers a systematic guide to planning and evaluating classroom assessments.

A PROCESS FOR PLANNING AND CONDUCTING CLASSROOM ASSESSMENTS

Assessment is an essential component of the teaching-learning process. Without effective classroom assessment, it is impossible for us to know whether our students are learning what we want them to learn. We have described briefly a framework of assessment approaches and meth-

ods available to teachers. In the remainder of this chapter we present a process for planning, administering, and interpreting results from classroom assessments. An Assessment Planning Chart assists in this process. We then illustrate the process in use.

Six Steps for Planning, Administering, and Interpreting Classroom Assessments

1. Identify the learning outcomes to be assessed, the purposes for the assessment, and the audiences for the assessment results.
2. Select methods from the Framework of Assessment Approaches in Figure 1 to assess the outcomes that are appropriate to your assessment purposes and audiences.
3. Record your decisions from steps one and two in an Assessment Planning Chart (see Figure 3). Evaluate the appropriateness of the:
   a. assessment approaches and methods you have chosen for your assessment outcomes, purposes, and audiences.

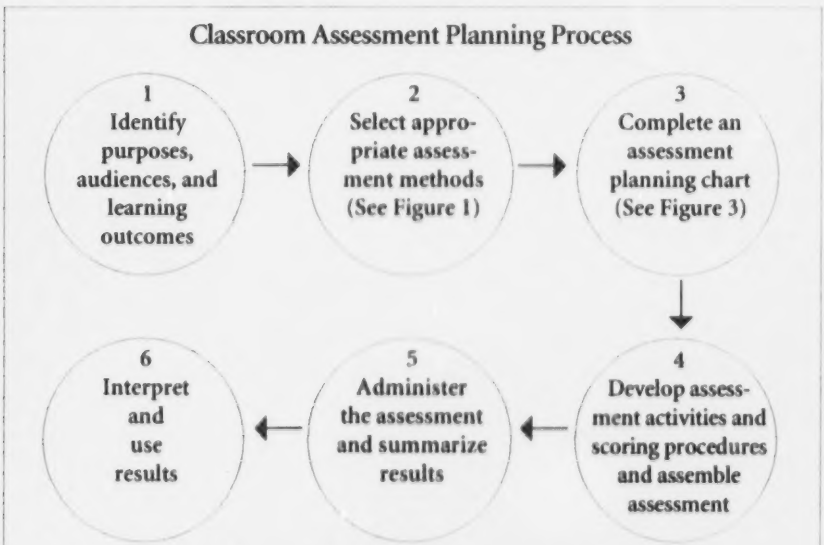


Figure 2: Six Steps for Planning, Administering, and Interpreting Classroom Assessments

   b. time allotted to each outcome and assessment activity and the time required to administer, score, and interpret all assessments.
4. Develop and assemble assessment activities and scoring procedures.
5. Administer the assessment and score, rate, or otherwise summarize the results of the assessment for your audiences.
6. Interpret and use the results, keeping in mind the learning outcomes you assessed and your assessment purposes and audiences. Consider potential consequences of your planned uses of the assessment results.[6]

In the next section we present an Assessment Planning Chart, a helpful tool for use in the classroom assessment planning process.

Assessment Planning Chart

Architects create blueprints to communicate their overall vision for a building and to specify details for a builder. The builder then constructs the architect's vision by executing the details in the blueprint. The result is a safe and long-lasting (and, hopefully, aesthetically pleasing) building. Similarly, classroom teachers can create assessment blueprints to communicate their overall vision for an assessment and to specify details such as outcomes to be assessed and assessment methods to be used. They then can use the blueprint to construct the envisioned assessment, developing activities that assess what has been specified. The result is twofold. First, the assessment that is developed from the blueprint produces valid information about what students know and can do. Second, the assessment clearly communicates to students what they should know and be able to do and what they actually know and can do.

An Assessment Planning Chart[7] is a guide to completing the first two steps in the classroom assessment planning process: identifying learning outcomes to be assessed and selecting appropriate assessment methods from the Framework of Assessment Approaches. The Assessment Planning Chart also helps us reinforce the

Assessment Planning Chart

Learning outcomes to be assessed: • Content • Process	ASSESSMENT METHODS				Number of assessment activities and amount of assessment time required or allowed
	Number of assessment activities and amount of assessment time required or allowed				TOTALS

Figure 3: Assessment Planning Chart

principles of multiple sources of data and the congruence of assessment and instruction discussed earlier. The generic Assessment Planning Chart (Figure 3) contains three elements necessary for planning thoughtful classroom assessments: content and process outcomes to be assessed, assessment methods to be used, and the amount of assessment time to be allotted to each outcome and method. Teachers can adapt the generic Assessment Planning Chart to their specific needs and situations, as we illustrate at the end of this chapter.

Using the Assessment Planning Chart

In this section we describe steps to complete an assessment planning chart. First, list each outcome to be assessed, one per row, specifying the desired content, skill or process, and habit of mind. Second, after referring to the Framework of Assessment Approaches, write in one or more assessment methods to assess each outcome, one per column. Next, determine the number of assessment activities appropriate for each content, process, habit, and assessment method. This is a matter of deciding the portion of the total assessment activities and time that should be allocated to each outcome using each assessment method. It is appropriate to write in numbers or percentages of assessment activities.

Space is provided at the end of each row and column to summarize the coverage of outcomes and use of each assessment method. Teachers can enter in each space the number of activities and amount of assessment time required or allowed to complete the assessment activities in each row and column. The same information can be entered in the space at the lower right-hand corner of the Assessment Planning Chart for a total assessment time. Teachers can refer to and modify each row total to clarify and control how much of any classroom assessment is devoted to each listed outcome, especially if each row total is expressed as a percentage of the total number of assessment activities or assessment time. Likewise, teachers can refer to and modify the column totals to clarify and control how much of the total assessment each method will comprise. Controlling the emphasis placed on each assessed outcome and the reliance on each assessment

method is important for thoughtful and sound classroom assessments and for accurately communicating to students what is important for them to know and be able to do.

While teachers generally do not question the usefulness of blueprints like the generic Assessment Planning Chart, they often express concern that such tools are too cumbersome and time consuming to use. We disagree. Experience has shown that, with practice, use of the assessment planning process and tools like the Assessment Planning Chart becomes automatic and efficient. When teachers internalize the process, planning and developing a classroom assessment become a matter of following the six steps, adapting the assessment planning chart to their curriculum, and selecting and applying the appropriate assessment methods. In fact, many teachers already use parts of the process as they plan and conduct assessments in their classrooms. An advantage of the planning process we describe here is its systematic nature; it specifies and clarifies the steps necessary for conducting sound classroom assessments. Similarly, the Assessment Planning Chart explicitly links desired learning outcomes with appropriate assessment methods. Finally, the Framework of Assessment Approaches suggests a wider array of assessment methods than teachers may currently consider and use. The Assessment Planning Chart can be particularly effective when teachers work together to develop "banks" of activities to assess common instructional outcomes. Use of such a systematic planning process by grade-level or department teams results in more coordinated, valid, and reliable classroom assessments.

An Illustration of the Classroom Assessment Planning Process

To demonstrate the classroom assessment planning process, using the Framework of Assessment Approaches and Assessment Planning Chart, we will refer to a hypothetical instructional unit for a middle school social studies class. Imagine a three-week-long instructional unit in which students explore the concepts of citizens' "rights and responsibilities" through an examination of various historical topics and contemporary issues. The intended

outcomes for this unit might be as stated below.
Students will:

- develop an understanding of the concepts and principles involved in determining citizens' rights and responsibilities.
- increase their knowledge and appreciation of the rights and responsibilities assured to United States citizens under the Bill of Rights.
- analyze contemporary issues involving citizens' rights and responsibilities by applying concepts and principles developed during the unit.

These learning outcomes stress concept development and application, or transfer, to new situations. Thus, the assessment methods used during the unit will reflect these instructional goals. Since the teacher and the students are the primary audiences for the assessment information, the teacher will use both formative and summative assessments during the unit. Information gained from the formative assessment activities will help the teacher determine if students are developing appropriate understandings of the concepts and principles so that he or she can adjust instruction as necessary and inform students about their progress. Summative assessment activities will be used similarly and as the basis for evaluating student performance in order to assign grades. By identifying purposes, audiences, and learning outcomes, Step 1 in the assessment planning process is complete.

Experience has shown that, with practice, use of the assessment planning process and tools like the Assessment Planning Chart becomes automatic and efficient.

Step 2 is completed by reviewing the methods in the Framework of Assessment Approaches and by selecting several assessment methods appropriate for the identified purposes, audiences, and learning outcomes. The teacher may decide that the summative assessments will include:

- *Written Test.* An in-class test near the end of the unit. Students will write a brief analysis of the rights and responsibilities contained within a specified constitutional amendment, on freedom of speech, for example.

- *Essay.* A persuasive essay that students will develop during the final days of the unit. They will state and defend their views on whether private citizens in the 1990s should be allowed to own firearms. Students will draft and refine these essays based on what they have learned in class, from outside reading and research, and from an in-class debate.
- *Debates.* Students will work on one of several teams to prepare for participation in one of several in-class debates. These debates will address several questions relevant to ownership of firearms: whether the need for U.S. citizens to own firearms for personal protection has increased or decreased since the late 18th century; whether citizens are likely to be more or less safe if the manufacture and sale of firearms were restricted; and whether an increase in safety assumed to result from such restrictions warrants denying citizens the right to bear arms guaranteed in the Bill of Rights.

In order to adjust instructional activities as necessary and to inform students throughout the unit regarding their progress on the learning outcomes for the unit, the teacher has planned several formative assessment activities. Some of these activities will assess students' understandings of the concepts of rights and responsibilities:

- *Written List.* Early in the unit, the teacher will have students work in small groups to discuss the rights guaranteed them as citizens of the United States. Following the discussion, students will individually prepare a written list of their constitutional rights as they understand them. Student responses to this open-ended assessment activity will give the teacher insight into students' understanding of important unit concepts.
- *Oral Questions.* Throughout the unit, the teacher will lead discussions and pose oral questions to the class to determine their understandings of, and ability to apply, the concepts and principles of citizens' rights and responsibilities.

Other formative assessment activities will focus on students' thinking strategies as they prepare for the debates:

- *Kid Watching.* As students work in teams to prepare for the debates, the teacher will listen and observe to determine if they are addressing relevant issues, employing relevant information, developing clear arguments, supporting their arguments, anticipating rebuttals, developing responses, and acknowledging different points of view.
- *Think-Aloud Activities.* Throughout the unit, the teacher will ask students to describe their reasoning processes by thinking out loud. By listening to students as they articulate their thoughts, the teacher can identify misconceptions and fallacious reasoning so as to provide needed clarification or assistance.

Completing Step 3 of the classroom assessment planning process will involve developing an Assessment Planning Chart that summarizes all of the assessment activities planned for the unit. An example is provided in Figure 4.

In this example, we have listed the three intended unit outcomes and the seven assessment activities planned for this unit. We have also indicated numbers of assessment activities and estimated the amount of time required for each activity. With this chart, the teacher has a clear picture of the activities and amounts of time devoted to assessing each outcome and the portion of all assessment activities each method comprises. The completed chart enables the teacher to evaluate the appropriateness and degree of balance between the learning outcomes and methods of assessment. As a result of this step, the teacher may decide to change the emphasis on one or another of the outcomes or to devote more or less time to one or more of the assessment activities. In addition to using the Assessment Planning Chart for planning assessments in an entire unit, the chart can be used to develop a single assessment, such as an end-of-unit test.

With the Assessment Planning Chart, the teacher has a clear picture of the activities and amounts of time devoted to assessing each outcome.

Completing Steps 4 through 6 in the process involves developing, administering, scoring, and interpreting the results from the in-class short-answer test, oral

ASSESSMENT PLANNING CHART for unit on "*rights and responsibilities.*"

OUTCOMES	ASSESSMENT METHODS: FORMATIVE ASSESSMENTS: List	Oral Questions	Kid Watching	"Think-Aloud"	SUMMATIVE ASSESSMENTS: Paper-and-Pencil Test	Essay	Debate	Summary of Assessment Activities
1 Knowledge	open-ended list of rights: 5-10 min. in groups, 5 min. individual work	5-10 questions estimated: 1 minute each	4-5 student groups 30 minutes		1 analysis 15-25 minutes			• all seven methods used
2 Understanding				5-10 students 5 minutes each		1 hour	3 debates 30 minutes each	• oral quest • kid watching • think-aloud • test • essay • debate
3 Application								• think-aloud • essay • debate
Estimated time for assessment activities	10-15 minutes	5-10 minutes	30 minutes	25-50 minutes	15-25 minutes	1 hour	1 1/2 hours	3 hrs. 55 min. to 4 hrs. 40 min.

Figure 4: Sample Assessment Planning Chart

discussion questions, and all other planned assessment activities. Step 4 is probably the most labor-intensive part of the classroom assessment planning process. In this step the teacher will develop, or select from an existing bank, assessment activities and accompanying scoring procedures and assemble the activities into a form for use with students. For example, the teacher might jot down notes for the listing activity, topics for oral questions, focuses for kid watching and think-aloud activities. But the teacher may want to provide the in-class test, persuasive essay prompt, and debate team assignments and topics to the students more formally and in writing.

Step 5 involves administering each assessment activity, scoring or otherwise evaluating student responses, and summarizing the qualities and characteristics of student responses. Using the "rights and responsibilities" unit to illustrate, the teacher may decide to evaluate students' responses in each summative assessment by giving one to two points for each written summary in the in-class test and writing comments in the margins regarding the comprehensiveness and accuracy of each summary; by rating the persuasive essay using a four- or six-point holistic scoring system that focuses on characteristics of persuasive writing and circling any grammatical errors; and by evaluating students' performances in the debates on several dimensions, such as their observance of rules of debating (which the teacher provided as part of the unit), oral discourse style, persuasiveness of arguments, and effectiveness in rebuttals. The teacher may decide to evaluate students' responses to the formative assessment less formally. For example, the teacher may provide one- or two-word comments on student lists of Constitutional rights, only oral comments on students' responses during oral questioning, and a short written note or brief conversation with students after the kid-watching and think-aloud activities. Decisions such as these are determined by the purposes and audiences for each assessment activity.

In the final step of this process, Step 6, the teacher will communicate interpretations, conclusions, decisions, and other actions to take based on the assessment results. As in Steps 1 through 5, the teacher will communicate in

ways that are appropriate for students, parents, administrators, teachers, and other audiences. For instance, the teacher may have adjusted instructional activities based on what he or she learned from the four formative assessment activities and told the students about these adjustments. In contrast, the teacher may inform students of their scores on the three summative assessments, perhaps indicating letter grades corresponding to each score, and incorporate these scores into the semester course grade. In order to communicate to students what is expected of them during the unit, the teacher would have informed students ahead of time how each assessment activity was to be evaluated and used. The information from all seven assessment activities could also be used for parent conferences and mid-semester progress reports of student performance and progress.

CLOSING

Effective assessment is part of effective teaching and learning (Schafer, 1991). In this chapter we have discussed principles of sound assessment, described a process for systematically planning classroom assessments, provided a framework for selecting assessment approaches and methods, and illustrated the use of the planning process. While we could have presented additional principles and practices of classroom assessment, we believe that teachers can go a long way toward developing more thoughtful classroom assessments by applying the ideas and tools in this chapter.

Authors' note: Our thanks to the Maryland Working Group on Assessing Thinking for their work on the Framework of Assessment Approaches, to Susan Ciotta and Sandi Benson for their fine work on the graphic displays, and to Barbara Reeves for her insightful comments on a draft of this chapter.

NOTES

1. Defined as tasks that are valued in their own right (Linn, Baker, & Dunbar, 1991) because they require students to directly apply valued skills, processes, and knowledge.
2. Thoughtful classroom assessments result from deliberate and systematic planning whereby teachers select methods of assessment that are congruent with their desired learning outcomes and their methods of instruction.
3. Classroom assessments can include paper-pencil tests, oral and written quizzes, homework assignments, class discussions, and

other classroom activities in which teachers gather information about student learning. While the terms *assessment, test, evaluation,* and *measurement* are often used interchangeably, we use the term *assessment,* as defined in this footnote, throughout this chapter.

4. The Framework of Assessment Approaches is based on the work of the Maryland Working Group on Assessing Thinking, whose members during 1987-1990 included state and local school system staff and university faculty in instruction and assessment.
5. However, we stress that these methods are not limited to only this purpose. Many writers and speakers have amply demonstrated the use of multiple-choice items to assess, for example, students' reasoning processes in science and mathematics as well as to determine if students can select a correct answer from among alternatives.
6. There can be negative or positive consequences to our use of assessment results. For example, placing too much faith in a multiple-choice test of science knowledge could mislead a teacher into believing that a student is adequately prepared to perform a laboratory experiment involving chemicals.
7. We have borrowed the term, Assessment Planning Chart, from Stiggins, Rubel, and Quellmalz (1988).

REFERENCES

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd ed.). New York: Holt, Rhinehart and Winston.

Ferrara, S., & McTighe, J. (in press). *The mindful school: How to assess in the thoughtful classroom.* Palatine, IL: Skylight Publishing.

Haertel, E., Ferrara, S., Korpi, M., & Prescott, B. (1984). *Testing in secondary schools: Student perspectives.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

McTighe, J., & Schollenberger, J. (1991). Why teach thinking: A statement of rationale. In A. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (rev. ed.)(pp. 2-5). Alexandria, VA: Association for Supervision and Curriculum Development.

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Testing in America.* Chestnut Hill, MA: Boston College.

Novak, J. D., & Gowin, B. D. (1984). *Learning how to learn.* New York: Cambridge University Press.

Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. *Educational Measure- ment: Issues and Practice, 10*(1), 3-6, 12.

Secretary's Commission on Achieving Necessary Skills. (1991). *What work requires of schools: A SCANS report for America 2000.* Washington, DC: U.S. Department of Labor.

Spady, W. G., & Marshall, K. J. (1991). Beyond traditional outcome-based education. *Educational Leader- ship, 49*(2), 67-75.

Skills in the Classroom (rev. ed.). Washington, D.C.: National Education Association.

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10*(1), 7-12.

Stiggins, R. J., Rubel, E., & Quellmalz, E. (1988). *Measuring thinking skills in the classroom* (rev. ed.). Washington, DC: National Education Association.

Journals and Portfolios

One of the most significant aspects is the opportunity to reflect on your portfolio with another thoughtful person.
–Bena Kallick, *If Minds Matter, Vol. II*

With the heavy emphasis on content coverage, test review, and job preparation, many educators have eliminated or neglected a powerful tool of learning. Students rarely have time to reflect on their own learning. Without the reflection piece in the learning photo album, students often do not take the time to process their ideas, learn from their mistakes, or celebrate their successes. Metacognition, the process of planning, monitoring, and evaluating one's own thinking, can be enhanced by having students keep reflective journals to record their feelings, reactions, elaborations, or contemplations. Sometimes a divided journal is effective because they can write their "Initial Reaction" on one side and then come back to the same page at a later date and write their ideas, "Upon Reflection." They can then see how their ideas change over time because of new developments or because of changing thought processes. Elaine Surbeck, Eunhye Park Han, and Joan E. Moyer recommend using three stages in journal writing: reaction, elaboration, and contemplation. They say that "Taking the time to write in a journal means also taking time to reflect."

Mary Renck Jalongo encourages teachers to chronicle their own teaching experiences in order to reflect upon what it means to teach and to learn. She says that people remember narrative stories and anecdotes, and "Stories and teaching enable us to organize, articulate, and communicate what we believe about teaching. . . ." Therefore,

teachers can become reflective practitioners who model the importance of processing and contemplation for their students.

Another authentic assessment tool that encourages students to evaluate their own work and to see their own development is the portfolio. Portfolios can be used in a wide variety of ways. Most importantly, they are used to help students take ownership of their own learning by giving them more control over the learning process and teaching them how to self-assess. Students can keep samples of "works in progress," writing samples, recordings of their oral readings or presentations, problem-solving examples, art work, logs, journals, conferences, interviews, pictures of projects, cassettes or videos of performances, observation checklists, or any variety of samples selected by the student and teacher to reflect the student's growth. Students can select their "best" work, or work that "still needs work."

The power of the portfolio is in the process of selection; the articulation of why the piece is good or why it still needs work; the self-evaluation process, and the ability to see growth and development in specific areas. Some schools encourage students to conduct their own portfolio conferences with the teacher or the parents so students learn how to verbalize their self-assessments. When students take ownership of their learning and can honestly assess themselves without relying on the teacher with the red pen or "happy face" stamp in hand to pass judgment, they learn a skill that will transfer to their academic, vocational, and personal lives. William Glasser advocates in his quality school programs that students should be critical of their own work and only put a "Q" for quality on the work they sincerely feel is quality work. If they take pride in their learning, it is their judgment, their self-assessment, and their self-evaluation that really counts. The portfolio also provides an opportunity for students to enlarge their "photo album" and include a richer picture of themselves.

Assessing Reflective Responses in Journals

by Elaine Surbeck, Eunhye Park Han, and Joan E. Moyer

In studying the journals of student teachers, three researchers categorized their responses in order to better understand the students' thinking and encourage them to become reflective practitioners.

"Why should I keep a journal?" We felt frustrated as we tried to answer our student's question with assurance. We knew that journal writing encourages thinking; we hoped it would even "serve the important function of integrating course content, self-knowledge, and practical experiences with teaching and learning situations" (Yinger and Clark 1981, p. 25). We also knew that journal entries can reveal how undergraduate teacher candidates think about teaching (see Zeichner and Liston 1987). Intrigued by the possibilities, we had added journals to our course requirements. What we did not know was how our students typically organize their thinking and whether there are common strategies by which their thinking could be examined for focus and content.

Taking the time to write in a journal means also taking time to reflect.

In their biweekly journal entries, we asked our students to focus on course content and readings as each related to (1) their beliefs and knowledge about teaching, (2) how such information was applicable in field experience and/or other classrooms, (3) how children responded (or might respond) to methods students experienced in class, and (4) personal reflections and feelings about the

From *Educational Leadership*, vol. 48, no. 6, p. 25-27, 1991. Reprinted with permission.

teaching/learning process, both in the college classroom and in the primary grades. They turned in their journals every month; then each professor read and responded to the entries.

Ten students volunteered to let us do an in-depth examination of their journals. Through naturalistic research, we developed the following framework of categories and subcategories of student responses from these journal entries.

REACTION

The Reaction category contained students' initial responses to class content, including peer teaching, discussions, activities, lectures, environments, instructors, peers, and articles they had read. The reactions included the following subcategories:

1. ***Positive Feeling:*** describing satisfaction or pleasure regarding themselves, others, or the class activities.

STUDENT #8 (August 29):

Mary [a guest speaker] *is very inspiring. Her enthusiasm for children's literature is exciting. I also share that excitement. I love the idea of incorporating literature and textbooks to give a creative, exciting curriculum.*

2. ***Negative Feeling:*** expressing dissatisfaction or complaints about themselves, others, or the class activities.

STUDENT #7 (October 24):

We gave our pre-number presentation. Our group endeavor seemed to be unsuccessful to me.

3. ***Report:*** stating facts of what the students did in class; does not include any personal concern or feelings about the class activities.

STUDENT #5 (September 19):

Toured science, math, and social studies labs.... Discussed Piagetian interview—directions for administering.... Also discussed, in our groups, thematic unit.

4. ***Personal Concern:*** expressing specific personal matters that had an impact on student's class participation.

STUDENT #7 (October 26):

I had a bad migraine headache. It lasted all week. I've had migraine headaches since I was 7. When they hit, they hit hard.

5. ***Issues:*** referring to educational issues, problems, knowledge, or related matters.

STUDENT #1 (August 24):

I think that parents, and even teachers and administrators, should be educated on the importance of developmental learning.

ELABORATION

Elaboration emerged as the second core category. The students expanded their first reactions by explaining their feelings, verifying their thinking, giving an example, or referring to other situations. Here we present examples of their first reactions and examples of the three types of elaboration.

The students expanded their first reactions by explaining their feelings, verifying their thinking, giving an example, or referring to other situations.

1. ***Concrete Elaboration:*** expanding on the particular situation from which the reaction was derived.

STUDENT #10 (August 29):

Reaction—*Today I learned how to go about making a list of what could possibly be used for a thematic unit.*

Concrete elaboration—*We brainstormed ideas, voted on one unit.... We also incorporated the holidays and special days that would come up during the year into our ideas for units for the year... .We webbed out or mapped out our thematic unit as a group.*

2. ***Comparative Elaboration:*** comparing their first reaction with other classes, primary classrooms or kindergartens they had been working in as field experience students, or their own childhood experiences.

STUDENT #6 (October 19):

Reaction—*Social studies and science appear to be so neglected in the educational system.*

Comparative elaboration—*The three semesters that I have been in an observation placement I have never observed science or social studies instruction. One teacher told me that none of the teachers knew science information, and they do not like the topics; therefore, they don't teach them.*

3. ***Generalized Elaboration:*** referring to a general principle, a theory, or a broad philosophical context.

STUDENT #3 (September 7):

Reaction—*Focusing on process and not so much on product is important to do in early childhood.*

Generalized elaboration—*Chapter 13 of* Piaget for the Classroom Teacher *made a good point about this on page 107. As quoted from Piaget himself, "Knowledge does not develop from 'all wrong' to 'perfectly correct.' All children have some knowledge about whatever we try to teach them. . . ."*

CONTEMPLATION

We called the last category contemplation. To be counted as contemplation, entries showing the initial reaction combined further elaboration, as well as thinking about personal, professional, or social/ethical problems.

1. ***The Personal Focus:*** thoughts about themselves or particular personal problems.

STUDENT #2 (September 21):

Reaction—*Today we had numerous guest speakers on the various cultures we are studying for our in-class units. The impact the speakers had on me was incredible.*

Elaboration—*I gained a greater appreciation for the varius cultures because I have a better understanding of them. I am very anxious for the reports to be shared in class.*

Contemplation—*I found myself going back in time viewing historical events from an entirely different perspective.*

2. ***The Professional Focus:*** thoughts about educational problems, issues, theories, teaching methods, future goals, attitudes as teachers, and views about children's abilities and nature.

STUDENT #9 (September 26):

Reaction—*Today was the day we started wrapping up our unit. The hardest part about this whole thing is trying to get together with each other.*

Elaboration—*Some of us live close, and some of us don't. It takes a lot of cooperation and giving and taking to be able to work together as a group.*

Contemplation—*I don't know if children would work as well together in such a large group. I feel that three at the most is large enough for a group activity.*

3. *The Social/Ethical Focus:* students' thoughts about social issues, ethical problems, or moral concerns.

STUDENT #7 (November 27):

Reaction—*I was really impressed by the graph that the professor made representing the population during different years. The prediction for the population for the year 2000 is frightening.*

Elaboration—*What kind of a world will my children live in? Are we the last generations? This changing and complex society is not the same as it was 30 years ago.*

Contemplation—*if our schools are going to abandon teaching religion—which I agree they must—and values, we have to replace it with something else. Children need direction. We need to help children be thinkers.... I've become more aware of the problems of society in the last 5 years. It seems so obvious that we are sinking fast. In a way I think that values have been exterminated by human rights.*

TOWARD A REFLECTIVE PERSPECTIVE

What did we the professors learn? We found that many students organized entries using the reaction-elaboration-contemplation sequence. That is, the sequence was at least partially developed in a majority of journals, though only a few entries included the contemplation category. When the complete sequence was present, we detected greater integration of information. The rough progression of thinking from personal and concrete reactions to a more sensitive social

I felt [the journal] was a wonderful way to expand my thinking of the events taking place in class.

and ethical perspective is a focus for future journal writing guidelines. But what about the student's initial query *Why should I keep a journal?* Let the students answer the question....

STUDENT #2 (December 5):

I felt [the journal] was a wonderful way to expand my thinking of the events taking place in class. It provided the challenge of questioning how I felt about concepts and how they could be applied in the classroom. Comprehension is greatly increased when students write information or concepts in their own words.

STUDENT #5 (December 5):

My journal also helped me to gain interesting insights that I might not have gained if I didn't write in my journal. Taking the time to write in a journal means also taking time to refelect.

Using journals this way takes time but assists prospective teachers in becoming better thinkers who probe deeper into both professional literature and their own teaching/learning ideas and actions. After all, "How teachers behave and what they do is directed in no small way by what they think" (Shulman and Lanier, cited in Fenstermacher 1978, p. 44).

REFERENCES

Fenstermacher, G. D. (1978). "A Philosophical Consideration of Recent Research on Teacher Effectiveness." *Review of Research in Special Education* 6, 1: 157-185.

Yinger, R. J., and M. Clark. (1981). *Reflective Journal Writing Theory and Practice* (Occasional Paper No. 50). East Lansing: Michigan State University, Institute for Research on Teaching.

Zeichner, K. M., and D. P. Liston. (1987). "Teaching Student Teachers to Reflect." *Harvard Educational Review* 57, 1: 23-48.

Using the Computer as a Reading Portfolio

by Judy Hetterscheidt, Lynn Pott, Kenneth Russell, and Jakke Tchang

In assessing their reading skills, 5th graders are taking ownership of their learning.

Standard measurements don't talk back nearly as well as kids do. With the help of a computer, we are assessing our 5th graders' reading progress by recording their voices as they read aloud, then saving the students' self-evaluations along with the recordings in a computer portfolio.

Bellerive Elementary, in suburban St. Louis, has been using a literature-based, process language approach for five years. Daily Writing and Reading Workshops in K-6 classes incorporate elements of both whole language and Donald Graves' process methods.[1]

Though pleased with the progress of our students in language arts, our 5th grade team nonetheless has experienced the teacher's old frustration: how do we pull students *inside* the learning process, so education isn't something that happens to them, but something they make happen?

Standard measurements don't talk back nearly as well as kids do.

Last year, Bellerive made portfolios a permanent part of each student's records. With the computer program *The Grady Profile—Portfolio Assessment*,[2] we scan and record two pieces of writing (one in September and one in June) and log the books each child reads during the year. The program also helps us maintain assessment checklists by subject and provides room for extensive teacher/student comments.

From *Educational Leadership*, vol. 48, no. 6, p. 73, 1992. Reprinted with permission.

With support from the Parkway School District and a Chapter II grant, we have extended the project to include all three 5th grade classes. Each teacher has a Macintosh computer, a microphone, and a printer.

Our goals are to involve students in self-evaluation and critical thinking, promote students' ownership of the learning process, and make reading "samples" part of student portfolios.

Today's 5th graders are wired for the electronic age. They have computers at home and associate them with challenge and skill.

As part of our Reading Workshops, we model and discuss good oral reading skills. At an initial reading conference, we explain that the computer can help students hear how they read and help them decide what to work on in reading. Each student chooses a brief passage to read: his or her own writing, published material, or a sample paragraph from *The Grady Profile*. After reviewing the passage silently, the student reads it into the microphone.

Before and after they listen to the playback, we talk about how well students think they read and where they can improve. We type their self-evaluations in *The Grady Profile*. As their summaries show, students can assess their performance accurately and select appropriate skills to work on:

> I stopped at periods and used a little expression. To be the best reader I need to stop stuttering at words. To accomplish this I need to look at the words more carefully. I read at home every night. I read to my mom and she reads to me. This helps me.

Reading conferences take 5-7 minutes per student. With the computer, everything is there: the microphone, reading sample, and assessment log. The student's reading is stored on the disk, which avoids having 30 separate audiotapes.

Today's 5th graders are wired for the electronic age. They have computers at home and associate them with challenge and skill. After a reading conference, a student said, "I've seen computers with a computer voice, but never one with *my* voice." A teacher isn't likely to find a juicier carrot to dangle in front of a reluctant learner.

When students hear their own words and make immediate assessments, they become discerning critics, aware of the learning *process* as well as the product. They can state *why* they do well, not "It's good," but "I lowered my voice at the end of sentences."

Being one-on-one with the teacher makes students feel special and provides teachers another chance to individualize instruction. The additional reading conference time extends the benefits of writing conferences because it focuses attention on a specific need and builds student confidence. We know our students better than we have ever known them.

Self-assessment gives students control. Recording their evaluation, not our judgment, transfers ownership to the rightful owner of learning...the student.

NOTES

1. D. Graves, (1983), *Writing: Teachers and Children at Work*, (Portsmouth, N.H.: Heinemann).
2. *The Grady Profile*—available from Aurbach and Associates, Inc., 8233 Tulane Ave., St. Louis, MO 63132—is a Macintosh, HyperCard application requiring two megabytes of internal memory and a hard drive.

Teachers' Stories: Our Ways of Knowing

by Mary Renck Jalongo

By sharing stories about their classroom experience, teachers not only gain insight into their own practice, but they also contribute to the storehouse of knowledge about teaching.

Consider for a moment how other professions use stories, not just as casual conversation, but as tools for professional growth. In medicine, stories are case histories; in law, stories set legal precedents; in business, real and hypothetical stories become scenarios. Experts from every walk of life organize their specialized knowledge and skill into episodes, events, or cases (Bruner 1988, Carter and Doyle 1989). Among the social sciences, education alone remains reluctant to share and value the stories that give form and meaning to our lives as educators. We need to use stories as other professional fields do, to treat stories as "little factories of understanding," using them to "attract and light up everything relevant" in our professional lives (Hughes 1988). Educators' stories about teaching and their reflections upon them are a deceptively simple way of addressing significant issues about what it means to teach and to learn (Frederick 1990, Schön 1983).

THE POWER OF THE NARRATIVE

Our technological society's insistence on "hard data," facts, and empirical research would tend to suggest that the story is an inferior way of knowing—narratives are "soft," subjective, and ungeneralizable. But there is nothing to be gained by creating artificial dichotomies between these two very different, yet complementary,

From *Educational Leadership*, vol. 49, no. 7, p. 68-73, 1992. Reprinted with permission.

ways of knowing (Abbs 1984, Bogdan 1980, Calderhead 1987). Almost to a person, the best people in our field have learned to allow their scientific and narrative modes to interact. As Sutton-Smith (1988) points out,

> The narrative mode has little to do with objectivity, predictors, and verifications; rather, it has to do with consensual support, impartial readings, and verisimilitude. The science that derived from physics and mathematics is a science of verification; the science that derives from linguistics and narratives is a science of interpretation (pp. 22-23).

AN ACT OF MIND

Stories are "a primary act of mind," a basic way of processing information (Hardy 1977). Both children and adults find it much easier to remember and use material presented in story format rather than as a categorized list (Bretherton 1984, Egan 1986). If you doubt this is true, attend an all-day workshop and consider what you recall in any great detail by the end of the day or much later. Chances are, it is a personal anecdote shared by a workshop leader or a participant.

It used to trouble me that whenever I encountered former students, they would invariably remark on my stories about children rather than the theories or research I was trying to convey. But now I have come around to Seymour Papert's (1990) perspective on learning:

> Understanding learning is my life-long passion.... But interestingly I find that what helps most is not the proliferation of abstract principles. I gain more by extending my collection of "learnings"—concrete learning situations that I can use as "objects to think with" (p. ix).

I now believe that the teaching stories I chose to share with students served their purpose. Those stories and our reflections upon them have been and continue to be "good things to think with."

THE STORY IS YOU

Susan Ohanian (1989) once observed, "The more I teach, the more I realize that we teachers are nothing but our

anecdotes, our reflections on experience." As teachers, we become the stories we choose to tell. If our personal narratives are primarily celebrations of student learning, we have high expectations for students; if the stories we choose to tell about teaching are little more than petty complaints, we have grown dull and apathetic; or, if our personal narratives are mainly tales of despair, we are "burned out" and in desperate need of renewal. This happens because personal narratives are a way of "structuring experience itself, laying down routes into memory, for not only guiding the life narrative up to the present but directing it into the future…a life as led is inseparable from a life as told…a life is not 'how it was' but how it is interpreted, told and retold" (Bruner 1988). Stories about teaching enable us to organize, articulate, and communicate what we believe about teaching and to reveal, in narrative style, what we have become as educators.

Stories about teaching enable us to organize, articulate, and communicate what we believe about teaching and to reveal, in narrative style, what we have become as educators.

CONTRIBUTIONS TO PROFESSIONAL GROWTH

Professional growth is more like finding our way through a forest than driving down a freeway; each of us must find our own path to professional fulfillment. Teacher stories contribute to that process of discovery.

Teachers' Stories Invite Reflection

One thing that differentiates reflective practice from routine practice is the number, richness, and flexibility of the "scripts" teachers bring to the classroom setting (Schön 1983, 1987, 1991).

As a teacher educator, I sometimes hear a veteran teacher complain that a novice teacher "just doesn't have common sense." Professional educators' common sense derives, not from rote memorization of many precise pieces of information, but from the stories used to make all those bits of information cohesive and relevant. Reflective practitioners have "common sense" precisely because they have a storehouse of stories that organize, apply, and

FIGURE 1: A Dozen Story Ideas*

1. **Reflections on Improvement** (Chism 1990)
   The change I made:
   What prompted the change:
   The impact of that change and
   How I assessed it:

2. **A Metaphor for Myself**
   What best symbolizes you as a teacher?
   What are the features that unify your self and the symbol you selected?

3. **At Least...**
   Generate a series of statements about making the best of a bad situation in your professional life (Our district may be in a budget crunch, but at least I ...)

4. **Joyful Moments**
   Make a list of the most joyful moments in your professional life.

5. **Imaginary Dialogue**
   Write an imaginary dialogue between your self and the school, an administrator, a colleague, a parent, or a child.

6. **Unrevealed Kindnesses**
   Write about a situation where you went above and beyond the call of duty to help a child in distress.

7. **Remembrances**
   Reflect upon your own experience as a child at school. Write about an incident from your childhood that has enabled you to develop greater empathy for the children you now teach.

8. **Low Points**
   Describe an incident that nearly caused you to abandon the teaching profession. Looking back on it, why was this such a critical incident?

9. **Heights and Depths**
   Write about your most and/or least successful learning experiences.

10. **Memorable Teachers**
    Profile the best and/or worst teacher you ever had.

11. **Packing Decisions**
    If you were invited to teach overseas and knew that your teaching resources would be very limited, what would you pack in your suitcase? Compare/contrast your choices with those of other teachers. What do your choices reveal about you as a teacher?

12. **Looking Back**
    Contact a former teacher you admired and interview him or her. What insights did you gain about the teacher? About yourself?

* Items 2-6 adapted from Cooper (1991).

interpret what they know about teaching (Shafer 1981). Figure 1 suggests specific strategies for using stories to encourage more reflective teaching.

Teachers' Stories Are a Metaphor for Change

Stories are not crystallized; they are fluid. As stories evolve, they sometimes seem to take on "a life of their own. New revelations of meaning open out of their images and patterns continually, stirred into reach by our own growth and changing circumstances" (Hughes 1988, p. 35).

As an illustration of the dynamic quality of narratives, consider this story that Krista, a preservice teacher, shared with a class of student teachers after her initial meeting with a small group of 1st graders:

> I was handing out construction paper and giving the children their choice of color when this child shouted, "I want black. Black is superior. Black is always superior."
>
> I thought, oh boy, I am really going to have problems with this one. I just never thought people had those racial attitudes so *young*.

One of the teachers in the group asked Krista if she would have made the same inferences about racial attitudes if the child had been white and had said, "I want white, white is always superior."

"Maybe not," Krista admitted, "maybe he just wants attention." Several weeks later, Krista shared with the class a different viewpoint about this child:

> Yesterday, I was teaching a lesson on the concept of celebrations. At the end, I asked the children to draw a picture and tell a story about a celebration they had experienced.
>
> This little boy said, "No! There ain't no celebrations at my house since my baby sister died."
>
> I found out that his sister had just died of leukemia. Next time, I won't be so quick to judge. I'll learn to focus more on the child and less on my own problems.

When I shared this story with a group of students, some were quick to judge Krista. Rose remarked, "I'll bet she felt ashamed." But when Rose had a personal experience with misinterpreting a child's behavior, she quickly recognized the similarities:

> On the first day of my summer school class, Joey came into my classroom, slumped in his chair, folded his arms across his chest, and mumbled obscenities. I took him aside and he stopped cursing, but he remained uncooperative for several days.
>
> Later, Joey confided that he knew he was going to have "a really mean" teacher next year in 3rd grade and that he had been afraid that summer school "would be the same." Evidently, the cursing was his way of defending himself.

Rose realized that "just as in the story you told about the boy whose sister had died, I was worrying about what *I* was going to do instead of wondering *why* is this kid behaving this way."

To see how stories become metaphors for change, notice how the same story was interpreted and reinterpreted. For Krista, what began as a "racial incident story" ended as a "becoming a more child-centered teacher story." For Rose, it was a story about a mistake she could not imagine herself committing. Yet later, she realized the issue was the same one she faced.

Personal narratives can reveal the nurturing dimension of the teaching role, characterize important changes in our professional lives, and encourage more reflective practice.

The same story not only encapsulates the transitions made by those directly involved but also transcends the boundaries of time and space when it is shared with others. That is because "a story is something happening to someone you have been led to care about… whatever its subject matter, every story is about change" (Shulevitz 1985, pp. 7, 47).

Teachers' Stories Promote the Ethic of Care

Asked about their concerns as preservice teachers, one student, Teri, responded, "Maybe this sounds funny—but

I worry about caring too much, about children's problems 'getting to me.'"

Teri's comments prompted me to share the following story about April, a kindergartner:

> The first child I noticed in the class was April. She looked neglected and seemed desperate for approval. I had seen April's teenaged mother pick her up at school once, but April's grandparents, who were openly resentful about having to care for her, appeared to be in charge.
>
> I was in the classroom one day as the children were getting ready to lie down on carpets for "quiet time." A classmate asked sarcastically, "Hey, April, where's your rug?" and another answered, "She probably doesn't even have one."
>
> April responded by making a funny face and dancing around wildly. When the derisive laughter faded, April walked over to the sink and pulled a handful of paper towels from the holder. Then she unfolded the paper towels, spread them out on the tile floor and curled up on top of them in an awkward fetal position. I couldn't stand it. I started to cry. April's teacher seemed oblivious to this dramatic Friday afternoon event.
>
> That weekend, I discussed the incident with my family. Almost before the story ended, my young nieces were rummaging through the linen closet. They thrust a small carpet into my hands. "Here. Take this to school and give it to her," they said. "If you're worried that the teacher will get mad, don't let her see you."

Teri responded to this story. She told me it "really helped me to feel it's okay to be sensitive and it's reasonable to take action." The story of April was a better response to Teri's concern than any other reason I could formulate. It was better because it demonstrated the ethic of care that must dominate our profession (Noddings 1984, Witherele and Noddings 1991).

When I shared the story with experienced teachers, they spoke—many of them for the first time in their professional lives—about "unrevealed kindness," things they had done for children without any expectations of reward or recognition. To me, that ethic of care and stories about it are like the mast of a ship on a turbulent sea; we lash ourselves to it as a defense against incessant

waves of change for change's sake and gales of criticism. Stories remind us of the reasons we went into teaching in the first place.

"ALL WE HAVE IS THE STORY"

Personal narratives can reveal the nurturing dimension of the teaching role, characterize important changes in our professional lives, and encourage more reflective practice. All of these benefits have a direct impact upon professional growth because "our lives are made of stories. Such stories allow us to explore our lives, to try out alternative possible ways of acting and being in the world, and indeed to help shape our future actions" (Kazemek 1985, p. 201).

Personal narratives are not superfluous features of teachers' lives; they are basic to our professional growth. Ultimately, "The world we know is the world we make in words, and all we have after years of work and struggle is the story" (Rouse 1978, p. 187).

REFERENCES

Abbs, P. (1984). *Autobiography in Education.* London: Heinemann.

Anderson, M. (1990). "Creating a Climate of Affirmation: Education Beyond Fear." In *Finding Our Own Way*, edited by J. Newman. Portsmouth, N.H.: Heinemann.

Bogdan, R. C. (February 1980). "The Soft Side of Hard Data." *Phi Delta Kappan* 61: 411-12.

Bretherton, I. (1984). *Symbolic Play.* New York: Academic Press.

Bruner, J. (October 1988). "Research Currents: Life as Narrative." *Language Arts* 65: 574-88.

Calderhead, J. (ed.). (1987). *Exploring Teacher's Thinking.* London: Cassell Educational Ltd.

Carter, K., and Doyle, W. (1989). "Classroom Research as a Resource for the Graduate Preparation of Teachers." In *Research Perspectives on the Graduate Preparation of Teachers*, edited by A. E. Woolfolk. Englewood Cliffs, N.J.: Prentice Hall.

Chism, N. (March 1989). "Reflective Teaching." Lecture delivered at Indiana University of Pennsylvania Indiana, Pa.

Cooper, J. E. (1991). "Telling Our Own Stories: The Reading and Writing of Journals or Diaries." In *Stories Lives Tell: Narrative and Dialogue in Education*, edited by C. Witherell and N. Noddings. New York: Teachers College Press.

Egan, K. (1986). *Teaching as Story Telling: An Alternative Approach to Teaching and Curriculum in the Elementary school.* Chicago, Ill.: University of Chicago Press.

Frederick, P. (December 1990). "The Power of Story." *AAHE Bulletin* 43, 4: 3.

Hardy, B. (1977). "Towards a Poetics of Fiction." In *The Cool Web,* edited by Margaret Meek et al. London: The Bodley Head.

Hughes, T. (1988). "Myth and Education." In *Imagination and Education,* edited by K. Egan and D. Nadaner. New York: Teachers College Press.

Kazemek, F. E. (December 1985). "Stories of Our Lives: Interviews and Oral Histories for Language Development." *Journal of Reading* 29: 211-18.

Noddings, N. (1984). *Caring: A Feminine Approach to Ethics and Moral Education.* Berkeley, Calif.: University of California Press.

Ohanian, S. (March 1989). Quoted in E. Schuster, "In Pursuit of Cultural Literacy." *Phi Delta Kappan* 70: 539-42.

Papert, S. (1990). "In Foreword to Upitis, Rena." *This Too Is Music.* Portsmouth, N.H.: Heinemann.

Rouse, J. (1978). *The Completed Gesture: Myth, Character, and Education.* New Jersey: Skyline Books.

Schon, D. A., ed. (1991). *The Reflective Turn: Case Studies on Educational Practice.* New York: Teachers College Press.

Schon, D. A. (1987). *Educating the Reflective Practitioner: How Professionals Think in Action.* New York: Basic Books.

Schon, D. A. (1983). *The Reflective Practitioners.* New York: Basic Books.

Shafer, R. (1981). "Narration in the Psychoanalytic Dialogue." In *On Narrative,* edited by W. T. J. Mitchell. Chicago: University of Chicago.

Shulevitz, U. (1985). *Writing with Pictures.* New York: Watson-Guoptill.

Sutton-Smith, B. (1988). "In Search of the Imagination." In *Imagination and Education,* edited by K. Egan and D. Nadaner. New York: Teachers College Press.

Witherele, C., & Noddings, N. (1991). *Stories Lives Tell: Narrative and Dialogue in Education.* New York: Teachers College Press.

Significant Outcomes

Our mission as educators is to help every child become a more active, engaged, committed, and skillful learner, not just for a test, but for a lifetime.–James Bellanca, *If Minds Matter, Vol. II*

What do you want your students to be able to do twenty-five years from now? That question helps define the importance of significant student outcomes. So often educators get bogged down in their attempts to teach discrete bits of knowledge, drill those skills incessantly, and, of course, test for those skills on Friday. Unfortunately, most students do not see the relevance of these skills and, therefore, "turn off" to school. William Glasser says students are not going to subject themselves to school if it does not satisfy their psychological needs for belonging, power, freedom, and fun.

Memorizing math formulas, parts of speech, spelling words, vocabulary words, the periodic table, the parts of a frog, and, of course, all the capitals of states and the presidents of the United States — in order — could be considered useful or even essential by some people. These discrete facts, however, may not serve students well twenty-five years from now. Many schools and school districts are moving beyond the "drill and skill" model of teaching to a broader category of significant student outcomes that will help students in their lives — not just on their tests. It is important to decide on the outcomes first and then to build the curriculum around those outcomes. *Every* objective, *every* lesson plan, *every* classroom activity, and *every* method of assessment should focus on helping students achieve those outcomes. Some

examples of significant outcomes selected by school districts include the following:

1. Communication
2. Problem solving
3. Decision making
4. Cooperation and collaboration
5. Technology
6. Self-esteem
7. Self-directed learning
8. Concern for others (sense of community)
9. Sense of responsibility
10. Higher-order thinking
11. Awareness of wellness and safety issues

If these outcomes are important enough for *all* students to know (not just the average or high-ability students), then it is important for teachers to ensure all students have achieved them. Tom Rusk Vickery states that "The instructional process does not prescribe or preclude any particular teaching strategy. Rather it is the teacher's professional responsibility to select from a wide range of instructional practices those that accommodate the learning styles of different students. . . ." Sometimes the information may have to be re-taught using a different teaching style. Sometimes students will have to work with others. Sometimes students will take longer to master those outcomes. Not all students will achieve the outcomes on the same day in the same way — but all students should achieve them if they are important for life.

The microcosm of discrete facts specific to one subject area is being replaced with the macrocosm of the big picture. No longer do we have the time to wait until all students "master" their basic skills *before* they move on to the more challenging higher-order thinking activities. Students may master those skills in the context of mastering the significant outcomes, but our primary focus needs to be on the life skills they need to transfer from the classroom to the community. No longer can we allow narrow behavioral objectives, out-of-date and out-of-touch curriculum guides, commercially made and standardized tests, and politically correct and watered-down

textbooks dictate what students need to learn. Educators need to re-evaluate their priorities and decide the key areas that teachers need to teach and students need to learn. If the outcomes are essential for students to achieve in order to have equal opportunity for a fulfilling life, then we cannot afford to provide that opportunity to only the "best and the brightest" of our children. All students should have the opportunity to become productive and contributing members of our society.

One of the most significant educational outcomes is higher-order thinking. Arthur L. Costa recommends that we must instill intelligent behaviors in our students and use observations and discussions to help measure thinking skills more authentically. In order to measure student growth accurately, we must "provide conditions conducive to the practice and demonstration of intelligent behavior."

Authentic assessment tasks recreate real-life situations that enable students to practice and to learn intelligent behaviors that go beyond the world of paper-and-pencil tests.

Learning From an Outcomes-Driven School District

by Tom Rusk Vickery

By reorganizing all aspects of schooling, the Johnson City, New York, schools have raised both standards and achievement in a community historically beset by low academic expectations.

In Johnson City (New York), the school system has systematically reorganized to achieve extraordinary student outcomes.[1] Students in Johnson City achieve at a level significantly higher than the modest expectations "appropriate" for a lower middle class community with virtually no managerial or professional citizens. Despite a free or reduced price lunch population of over 20 percent, a sizable Asian immigrant population with limited English proficiency, and the second highest poverty rate of the 10 urban districts in Broome County, district students steadily increase the margin by which they exceed the national norms on standardized tests. By the time students complete the 8th month of the 8th grade, they score, on average, at grade 11 or above in both reading and mathematics. By the time they finish the 8th grade, at least 75 percent of the students are at least six months above grade level.

Laudable test scores such as these, however, do not adequately reflect the district's goals. As its mission statement says, Johnson City's goal is for all students to learn well what the schools want them to learn. Many districts could subscribe to that mission statement and would measure the achievement of that learning primarily, if not exclusively, through standardized tests. But in Johnson City five exit behaviors further define this mission:

From *Educational Leadership*, vol. 45, no. 2, p. 52-56, 1988.
Reprinted with permission.

1. Students are to have high self-esteem both as learners and as persons.
2. They will be able to function at high cognitive levels, not just at the lower level expected on standardized tests.
3. They will be good problem solvers, communicators, and decision makers; will be competent in group processes; and will be accountable for their own behavior.
4. They will be self-directed learners.
5. They will have concern for others.

MASTERY LEARNING-BASED INSTRUCTION

The restructuring which is the point of this article is most obvious when one looks at the instructional program. Over time Johnson City has developed a clear and consistent definition of the instructional process. Based on mastery learning, that process ensures, first, that there is a high degree of correspondence between instructional objectives, what is taught, and what is tested; and, second, that more time for learning and more instruction are available for those who need it.

Despite a free or reduced price lunch population of over 20 percent, a sizable Asian immigrant population with limited English proficiency, and the second highest poverty rate of the 10 urban districts in Broome County, district students steadily increase the margin by which they exceed the national norms on standardized tests.

The district begins by assessing whether students have the prerequisites for the unit. If they do not, then the teacher provides instruction on the prerequisites. Then comes cue setting, a brief explanation of the unit's objectives, and a description of what students will be able to do and will know after mastering the unit. Next comes the "best shot" instruction; this is large-group instruction that the teacher thinks has the best chance of enabling all the students to achieve mastery. Following this instruction is guided practice in which the teacher informally evaluates how well each student is doing. The emphasis is always on successful practice; no child, for example, is asked to do homework that he or she does not know how to do.

When the teacher is confident that most, if not all, students can demonstrate mastery, he or she administers a formative test on the unit's objectives. Students who have mastered the objectives then work on enrichment activities, while the others receive corrective instruction. Thus, students who need more time and instruction receive it, and those who do not are freed to work on other things. Before the teacher goes on to the next unit, students take a summative test on the current one.

Instead of being graded on each unit, students are certified either as having mastered the unit's objectives or as not having completed the unit. The latter students are responsible for completing the unit at a mastery level after school or at other available times before the end of the course. Even though the class moves on to another unit, the teacher continues to work with students who have "incompletes" on previous units. Course grades are given every 10 weeks. The lowest level of achievement for which students can receive credit for a unit is 80 or 85, depending on the subject and level. If a student scores lower than that, he or she gets an incomplete, not a failing grade.

RESTRUCTURED RESPONSIBILITIES

An important part of the teaching-learning process is the restructuring of responsibility. Students, for instance, may not take a second summative test on a unit until they can prove that they have engaged in additional learning activities designed to help them achieve the unit's objectives. A student who has not done his or her homework hardly qualifies for extra consideration; rather it is up to the student to prove that he or she deserves corrective instruction by cooperating in the instructional process.

Not only does this procedure reduce demands on teacher time, but it also reinforces the district's stated goal that students will become self-directed learners accountable for their own behavior. Students know that if they assume responsibility for their part of the teaching-learning process, then they will receive additional time and instructional help. When students do not achieve—because of a lack of cooperation and effort, poor attendance, or a discipline problem—a districtwide policy

based on William Glasser's Reality Therapy is uniformly applied to help them take responsibility for their own behavior.

The instructional process does not prescribe or preclude any particular teaching strategy. Rather it is the teacher's professional responsibility to select from a wide range of instructional practices, those that accommodate the learning styles of different students: lecture, discussion and interaction, reflecting on and analyzing experience, integrating reflective analysis into concepts, practicing clearly defined concepts and skills, adding something of oneself to that which is being studied, and so on. Here is where staff development programs in—for example—Reading in the Content Areas, Cooperative Learning Groups, and Reality Therapy are most important: they expand the teacher's repertoire of instructional tools.

Efforts to achieve the district's exit behaviors have led to a restructuring of a total curriculum. Programs such as Talents Unlimited increase student creativity and self-esteem. Other programs help all students learn to pursue through independent investigations those higher cognitive goals that students have identified as important. Through Cooperative Learning Groups, for example, students gain process skills, and through volunteer community activities they learn to express their concern for others. For example, two years ago, the district's 2,700 students raised $15,000 to help the town's senior citizens build a center. And last year at Thanksgiving, the students collected more food for the needy than did the students from all 11 other school districts in the county.

Every aspect of school life is subject to change *if* that change increases the probability of achieving desired outcomes.

INSTRUCTIONAL LEADERSHIP IS CENTRAL

Leadership in the district is first and foremost *instructional* leadership. Every aspect of school life is subject to change *if* that change increases the probability of achieving the district's desired outcomes. It is not enough to introduce a new instructional process; any innovation must be sup-

ported by a curriculum, by school practices, and by organizational structures, all intentionally aligned toward achieving the same outcomes. And that intentional alignment does not stop with school support systems but extends to the whole administrative system and to the board of education. School board policy requires all decisions to be based on the best research literature available. Central administration is responsible for seeing that this intentional alignment of school practices with each other and with the research literature extends to staff development; to teacher evaluation; and to efforts to improve climate, solve problems, and manage the change and communications process.

For example, the attendance policy should support the instructional program; certainly it must not work against it by allowing a student to escape responsibility for his or her own learning. And the same is true for the system by which students are advanced from grade to grade. School board policies must be consistent with these efforts. Conscious attempts are made to develop and use models of change and communication that ensure that nothing is left to happenstance. Here and elsewhere instructional leadership from the superintendent and consequently from the rest of the administrative staff clearly is central to the district's success in promoting excellence in learning.

Typically, teachers are inundated with reading material about problems or practices. Then, loaded with new information on theory and research, they are given risk-free opportunities to develop programs consistent with that research and with district goals. Someone—a teacher or an administrator—will identify an instructional practice that promises to solve an existing problem or to improve achievement in some way. If it is consistent with the beliefs and practices of the district, then several people will learn all they can about it, usually by going to a training program at district expense. Then, with better knowledge of the practice, district staff will evaluate it in terms of its appropriateness for the district. If it is deemed appropriate, then the person or persons trained in that practice—called *core people*—make a presentation to the

district's staff inviting anyone interested to join them in its implementation. Core people thus become in-house trainers, gaining both expertise and a strong sense of ownership.

STAFF DEVELOPMENT MASTERY MODEL

Johnson City School District also operates on a mastery model in staff development. Thus, innovations are not implemented districtwide simultaneously, rather, staff members learn and deploy at their own rates. The rule is "You don't have to cooperate, but don't get in the way."

The district has a clear procedure for career development. As a learner, a staff member first must master the facts about innovation being implemented. Then comes application of the practice in controlled and predictable situations, as in implementing a unit guide prepared by someone else. Third, the teacher applies the practice in additional areas. And fourth, the teacher designs new courses or new units based on the practice.

As a teacher of teachers, he or she must first know the theory underlying the practice. Next comes modeling the practice for others. The third step involves the planning, say, of unit guides to help others learn to apply the practice. Finally comes the coaching of other teachers. Teachers whose proposals to assume leadership in solving a district problem are accepted receive released time and any other resources needed to enable them to function in their new roles.

Although all teachers in Johnson City are on display every year when hundreds of educators visit the district, a second aspect of the staff development program places a number of teachers and administrators even more in the public eye. They are encouraged to become district spokespersons, serving as consultants to other districts that wish to learn about and perhaps adopt an outcomes-driven mastery model. All staff are encouraged to become experts in the instructional process; those who serve as consultants are those who, in addition to gaining expertise, also become adept at explaining the program and at helping others try it out. Although becoming consultants gives teachers and administrators motivation to become experts at what they do and provides them with both

status and financial rewards, it costs Johnson City nothing. Any school employing the staff member pays not only a consultant fee but also for the substitute replacing the staff member while he or she is a consultant.

OVERCOMING HISTORICAL OBSTACLES

All this is not to say that the Johnson City School District has achieved perfection. Formidable problems still must be overcome. Johnson City bears the imprint of George F. Johnson, once head of the Endicott-Johnson Shoe Company, whose family name graces the town, the school district, and three of the four schools. Until his factory closed in 1960, George Johnson used to come to the high school to dissuade students from any aspirations for college; if his shoe factories were good enough for their parents, he told them, they were good enough for them. Johnson was a pioneer in providing medical benefits, parks, and subsidized housing and food for his workers, but he had no use for unions or colleges.

Students' aspiration levels are rising, as reflected in the number who take advanced placement courses and college admissions tests and in the number who win New York State Regents' scholarships.

With such a history, the district has faced the challenge of finding ways to see that high school graduates have both the competencies and the credentials to go to college, whether or not they do so. Staff members must influence the values of both the community and the students so that their graduates will attain equality both in their post-secondary aspirations and in their achievements. In this long and difficult process, results are already being seen. Almost 70 percent of the district's graduates earn the more difficult Regents' diploma necessary for admission to many campuses of the State University of New York, compared to a statewide average of less than 50 percent. That record is comparable to two of the three upper-middle class districts in the county and about 15 percentage points higher than the third.

Students' aspiration levels are rising, as reflected in the number who take advanced placement courses and

college admissions tests and in the number who win New York State Regents' scholarships. The patterns of college attendance are also changing. More students are attending four-year institutions, and a greater number are sending transcripts to MIT, the University of Pennsylvania, Harvard, Stanford, and other high-status schools.

PURSUING EXCELLENCE

Johnson City is a shining example of a school district that has been systematically reorganized to raise both standards and achievement. Leadership has become, first and foremost, instructional leadership, and staff development has become a means for sharing both leadership and responsibility for solving problems. It is a lesson in what can be done when a district commits every aspect of schooling to the pursuit of excellence.

Author's note: The following research reports are available from me: "Excellence in an Outcomes-Driven School District, A Validation Study of the Schools of Johnson City, New York," September 1985; "Changes in Variance Under Mastery Learning" (with Maritza Moreno de Suarez), report presented at the Annual Meeting of AERA, Chicago, 1985; and "Evaluating a Mastery learning High School," report presented at the Annual Meeting of AERA, Washington, D.C., 1987.

NOTES

1. Johnson City's program has been recognized by the National Diffusion Network for its excellence and for the Outcomes-Driven Development Model the district created to restructure all aspects of district life in its objectives.

Thinking: How Do We Know Students Are Getting Better At It?

by Arthur L. Costa

Educators have the great responsibility of instilling intelligent behavior in students, to teach them to value intelligence and rational action. They must provide conditioning conducive to the practice and demonstration of intelligent behavior, and believe that all students can continue to grow in their ability to behave more intelligently.

When considering how evidence of students' achievement is collected, we most often think of testing—using some form of paper and pencil instrument to determine how many questions a student answers correctly. While there may be some types of thinking that can be assessed in this fashion, we must seek additional means of determining growth in intellectual abilities. In teaching students to think the emphasis is not on how many answers they know. Rather, the focus is on how they behave when they don't know. We are interested in observing how students produce knowledge rather than how they merely reproduce knowledge. A critical characteristic of intellectual ability is not only having information, but knowing how to act upon it.

In teaching students to think the emphasis is not on how many answers they know. Rather, the focus is on how they behave when they don't know.

By definition, a problem is any stimulus, question, or phenomenon, the explanation for which is not immediately known. Thus, student performance is assessed under challenging conditions which demand strategies, insightfulness, perseverance, and craftsmanship to resolve the problem.

From *Roeper Review*, vol. 6, no. 4, p. 197-199, April 1984. Reprinted with permission.

Many teachers are unimpressed with standardized tests (Harootunian and Yarger, 1980; Lazar-Morrison, et al., 1980). The results are often unavailable for several weeks or months after administering the test; thinking skills are contaminated by the degree to which students are acquainted with the subject matter; behavior is influenced by the mental and emotional state of the students at the time of testing; performance is subject to the vicissitudes of the situation; scores yield neither diagnostic clues as to how the student derived the answer (meta-cognition) nor information on how the student processed the data and emotions necessary to arrive at the best answer (cognitive mapping) (Anderson, 1981; Coffman, 1980).

Teachers know that many students are deemed gifted because they are test wise. They also know that other students are often overlooked because they do poorly under testing conditions. While competency may be demonstrated in a single test, effectiveness is demonstrated by sustained performance in a variety of situations which demand the selective and spontaneous use of clusters and linkages of problem-solving strategies rather than singular, isolated behaviors.

As students interact with real life and day-to-day classroom problems, what might teachers search for as indicators that their instructional efforts are paying off? Following are ten suggested characteristics of intellectual growth which teachers can observe and record (Feuerstein, 1980). Keeping anecdotal records of a student's acquisition of these types of behaviors provides more usable information about growth in intellectual behaviors than typical norm-referenced, multiple choice, standardized achievement tests.

1. PERSEVERING WHEN THE SOLUTION TO A PROBLEM IS NOT IMMEDIATELY APPARENT

Students often give up in despair when the answer to a problem is not immediately known. We can observe them crumple their papers and throw them away saying, "I can't," or "It's too hard." They lack the ability to analyze a problem, to develop a system, structure, or strategy of problem attack.

Thinking students grow in their ability to use alternative strategies of problem solving. If one strategy doesn't work, they know how to back up and try another. They realize that the theory or idea must be rejected and another employed. They have systematic methods of analyzing a problem, knowing ways to begin, knowing what steps must be performed, what data need to be generated or collected. This is what is meant by perseverance.

2. DECREASING IMPULSIVITY

Often students blurt the first answer that comes to mind. Sometimes they shout out an answer, start to work without fully understanding the directions, lack a plan or strategy for approaching a problem, or make immediate value judgments about an aide before fully understanding it.

As students become less impulsive, we can observe them decreasing the number of erasures on their papers, gathering much information before they begin a task, taking time to reflect on an answer before giving it, making sure they understand directions before beginning a task, listening to alternative points of view, and planning a strategy for solving a problem.

3. FLEXIBILITY IN THINKING

Some students have difficulty in considering alternative points of view, or dealing with more than one classification system simultaneously. Their way to solve a problem seems to be the only way. They are more interested in knowing whether their answer is correct or not, rather than being challenged by the process of finding the answer. They are unable to sustain a process of problem solving and finding the answer over time, and therefore avoid ambiguous situations. They have a need for certainty rather than an inclination for doubt.

As students become more flexible in their thinking they can be heard considering, expressing, or paraphrasing another person's point of view or rationale. They can state several ways of solving the same problem and can evaluate the merits and consequences of two or more alternate courses of action. When making decisions they will often

use such words as "however," "on the other hand," "if you look at it another way . . ." or "John's idea is . . ., but Mary's idea is . . . ."

4. META-COGNITION: THE ABILITY TO KNOW WHAT WE DON'T KNOW

Some people are unaware of their own thinking processes. They are unable to describe the steps or strategies they use during the act of problem solving. They cannot transform into words the visual images held in their minds. They seldom evaluate the quality of their own thinking skills.

We can determine if students are becoming more aware of their own thinking as they are able to describe what goes on in their head when they think. When asked, they can list the steps and tell where they are in the sequence of a problem-solving strategy. They can trace the pathways and blind alleys they took on the road to a problem solution. They can describe what data are lacking and their plans for producing those data (Sternberg and Wagner, 1982).

5. CHECKING FOR ACCURACY

Students are often careless when turning in their completed work. When asked if they have checked over their papers, they might say, "No, I'm done." They seem to feel little inclination to reflect upon the accuracy of their work, to contemplate their precision, or to take pride in their accomplishments. Speed of getting the assignment over with surpasses their desire for craftsmanship.

We can observe students growing in their desire for accuracy as they take time to check over their tests and papers, as they grow more conscientious about precision, clarity, and perfection. They go back over the rules by which they were to abide, the models they were to follow, and the criteria they were to employ to confirm that their finished product matches exactly.

6. PROBLEM POSING

One of the distinguishing characteristics between humans and other forms of life is our ability to find problems. Yet, students depend on others to solve problems, to find answers, and to ask questions for them (Brown, 1983).

Over time, we want to see a shift from teachers asking questions and posing problems, toward students asking questions and finding problems for themselves. Furthermore, the types of questions students ask should change and become more specific and profound. For example:

We seek an increase in requests for data to support conclusions and assumptions. Such questions as, "What evidence do you have . . .?" or "How do you know that's true?" will be heard.

We want to hear more hypothetical problems. These are characterized by "iffy" type questions: "What do you think would happen if . . .," or "If that is true, then . . .?"

We want students to recognize discrepancies and phenomena in their environment and to inquire into their causes: "Why do cats purr?" "How high can birds fly?" "Why does the hair on my head grow so fast, but the hair on my arms and legs doesn't?" "What would happen if we put the salt water fish in the fresh water aquarium?" "What are some alternative solutions to international conflicts other than wars?"

7. DRAWING ON PAST KNOWLEDGE AND EXPERIENCES

Too often students begin each new task as if it were being approached for the very first time. Teachers are often dismayed when they invite students to recall how they solved a similar problem previously and students don't remember. It's as if they never heard of it before, even though they had the same type of problem just recently. It is as if each experience is encapsulated and has no relationship to what has come before or what comes afterward.

Thinking students are able to abstract meaning from one experience and carry it forth to apply it in the next experience. Students can be observed growing in this ability as they are heard saying "This reminds me of . . . " or "This is just like the time when we . . . " They explain what they are doing now in terms of analogies with or references to previous experiences. They call upon their store of knowledge and experience as sources of data to

support, theories to explain, or processes to solve each new challenge.

8. TRANSFERENCE BEYOND THE LEARNING SITUATION

Probably the ultimate goal of teaching thinking skills is for the students to apply school-learned knowledge to real-life situations and to content areas beyond that in which it was taught. Yet, we find that while students can pass mastery tests in mathematics, for example, they often have difficulty in deciding whether to buy six items for $2.38 or seven for $2.86 at the supermarket.

When parents and other teachers report how they have observed students thinking at home or in other classes, we know students are transferring.

When parents and other teachers report how they have observed students thinking at home or in other classes, we know students are transferring. Parents, for example, may report increased interest in school, more planning in their child's use of time and finances, in increased organization of their room, their books, and their desks at home.

We might hear, for example, the social studies teacher describe how a student used a problem-solving strategy which was originally learned in the science class. We might hear the wood shop teacher tell how a student volunteered to plan to measure accurately before cutting a piece of wood: "measure twice and cut once," an axiom learned in the math class.

9. PRECISION OF LANGUAGE

Some students' language is confused, vague, or imprecise. They describe attributes of objects or events with such non-specific words as "weird," "nice," or "OK." Names for objects are such as "stuff," "junk," or "things." Their sentences are often punctuated with "ya know," "er," and "uh."

As students' language becomes more precise, they can be heard using more descriptive words to distinguish attributes. They will use more correct names, and when universal labels are unavailable, they will use analogies

such as "crescent shaped" or "like a bow tie." They will speak in complete sentences, clarify, and operationally define their terminology. Their speech will become more concise, descriptive, and coherent.

10. ENJOYMENT OF PROBLEM SOLVING

Some children and adults avoid problems. We often hear them saying something like, "These types of thinking games turn me off," "I was never good at these brain teasers," or "Go ask your father; he's the brain in this family." Many people never enrolled in another math class or other "hard" academic subjects after they didn't have to in high school or college. Many people perceive thinking as hard work and therefore recoil from situations which demand "too much" of it.

We want to observe students moving not only toward an "I *can*" attitude, but also toward an "I *enjoy*" feeling. We will notice students seeking out problems to solve on their own and submit to others. They will make up problems to solve on their own and request them from others. Furthermore, students will solve problems with increasing independence, without the teacher's help or intervention. Such statements as "Don't tell me the answer, I can figure it out by myself," will indicate growing autonomy.

CONCLUSION

Assessing thinking through exclusive use of standardized, paper-pencil techniques is inadequate. Each teacher should look for indicators of growing intelligence. Most teachers neither need nor use standardized tests to determine the effectiveness of their teaching for thinking. Significant problem-solving behaviors are displayed and can be observed daily if we know how to recognize them.

As educators, we have the great responsibility of instilling intelligent behaviors in our students. We must teach them to value intelligent and rational action. To do so, we must provide conditions conducive to the practice and demonstration of intelligent behavior. We must believe that all students can continue to grow in their ability to behave more intelligently, and we must have

faith in the ability of all humans to become increasingly more gifted. Finally, we must set an example by modeling these intelligent behaviors ourselves (Costa, 1983).

REFERENCES

Anderson, A., "Testing and Coaching." Paper presented at the annual meeting of the American Association of School Administrators, Atlanta, Georgia, 1981.

Brown, I., *The Art of Problem Posing.* The Franklin Institute Press, Philadelphia, Pennsylvania, 1983.

Coffman, W. E. "Those Achievement Tests—How Useful?" in *Executive Review,* Vol. 1, No. 1, 1980.

Costa, A. "Teaching Toward Intelligent Behavior" in *Thinking The Expanding Frontier.* W. Maxwell, Editor. Franklin Institute Press, Philadelphia, Pennsylvania, 1983.

Feuerstein, R., *Instrumental Enrichment.* Teachers who have taught this program report changes in student behaviors similar to those described herein.

Harootunian, B. and Yarger, D., *Teachers' Conceptions Of Their Own Success.* Eric Clearinghouse on Teacher Education, Washington, DC, No. SP 017372, 1981.

Lazar-Morrison, C. and others, *A Review of the Literature on Test Use.* Center for the Study of Education, Los Angeles, 1980.

Sternberg, R. and Wagner, R., "Understanding Intelligence: What's In It for Educators?" Paper presented to the National Commission on Excellence in Education, July, 1980.

Authors

Rexford Brown is Communications Director and Director of a project called "Policy and the Higher Literacies" at the Education Commission of the States (ECS) in Denver, Colorado.

Arthur L. Costa is professor emeritus of Educational Administration at California State University, Sacramento and past president of the Association for Supervision and Curriculum Development.

Stephen B. Dunbar is an associate professor of measurement and statistics at the University of Iowa, Iowa City.

Steven Ferrara is director of student assessment for the Maryland State Department of Education, and the National Council on Measurement in Education's at-large representative to the Joint Committee on Testing Practices.

Eunhye Park Han is a doctoral candidate at Arizona State University.

Walter Haney is an associate professor of education and director of the Educational Technology Program at Boston College.

Judy Hetterscheidt is a member of the 5th grade teaching team at Bellerive School, Creve Coeur, Missouri.

Mary Renck Jalongo is Professor of Education, in the department of Professional Studies in Education, at Indiana University of Pennsylvania, Indiana, Pennsylvania.

Mara Krechevsky is Project Director, Project Spectrum, Harvard Project Zero, Harvard Graduate School of Education, Cambridge, Massachusetts.

Robert L. Linn is co-director of the Center for Research on Evaluation, Standards, and Student Testing and a professor of education at the University of Colorado, Boulder.

George Madaus is a professor of education and director of the Study of Testing, Evaluation, and Education Policy.

Carolyn Mamchur is an associate professor in the Faculty of Education at Simon Fraser University, Burnaby, B.C.

Jay McTighe works with the Maryland State Department of Education where he administers a statewide school improvement program.

Joan E. Moyer is a Professor at Arizona State University.

Fred M. Newmann is a professor of curriculum and instruction at the University of Wisconsin, Madison, and director of the Center on Organization and Restructuring of Schools.

Desmond L. Nuttall is Director of the Centre for Educational Research, London School of Economics and Political Science, and Professor of Curriculum and Assessment Studies at the University of London Institute of Education, London, England.

Chris Pipho is director of the Information Clearinghouse, Education Commission of the States, Denver.

Lynn Pott is a member of the 5th grade teaching team at Bellerive School, Creve Coeur, Missouri.

Kenneth Russell is Principal at Bellerive School, Creve Coeur, Missouri.

Lorrie A. Shepard is Professor, University of Colorado at Boulder, School of Education, Boulder, Colorado.

Richard J. Stiggins is director of the Center for Performance Assessment, Northwest Regional Educational Laboratory, Portland, Oregon.

Elaine Surbeck is Assistant Professor, Arizona State University, College of Education, Tempe, Arizona.

Jakke Tchang is a member of the 5th grade teaching team at Bellerive School, Creve Coeur, Missouri.

Richard VanDeWeghe is English Department Chair, University of Colorado at Denver. He is also Director of the Higher Literacy Project, Douglas County Public School District.

Tom Rusk Vickery is Associate Professor, Cultural Foundations and Curriculum, Syracuse University, School of Education.

Grant Wiggins is a special consultant to the National Center on Education and the Economy in Rochester, New York, and the Coalition of Essential Schools on issues of assessment and curriculum.

Acknowledgments

Grateful acknowledgment is made to the following authors and agents for their permission to reprint copyrighted materials:

Roeper Review for "Thinking: How Do We Know Students Are Getting Better At It?" by Arthur L. Costa. Vol. 6, no. 4, p. 197-99, April 1984. Reprinted with permission of *Roeper Review*, PO Box 329, Bloomfield Hills, MI 48303. Copyright 1984. All rights reserved.

Skylight Publishing for "Assessment: A Thoughtful Process" by Steven Ferrara and Jay McTighe. From *If Minds Matter: A Foreword to the Future, Vol. II*, Arthur L. Costa, James Bellanca, and Robin Fogarty, eds., 1992, p. 337-48. Reprinted with permission of Skylight Publishing, 200 E. Wood St., Palatine, IL 60067. Copyright 1992. All rights reserved.

The Association for Supervision and Curriculum Development for "Why We Need Better Assessments" by Lorrie A. Shepard. From *Educational Leadership*, vol. 46, no. 7, p. 4-9, 1989. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1989 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "Testing and Thoughtfulness" by Rexford Brown. From *Educational Leadership*, vol. 46, no. 7, p. 31-33, 1989. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1989 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "Teaching to the (Authentic) Test" by Grant Wiggins. From *Educational Leadership*, vol. 46, no. 7, p. 41-47, 1989. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1989 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "Project Spectrum: An Innovative Assessment Alternative" by Mara Krechevsky. From *Educational Leadership*, vol. 48, no. 5, p. 43-48, 1991. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1991 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "Performance Assessment: The Message from England" by Desmond L. Nuttall. From *Educational Leadership*, vol. 49, no. 8, p. 54-57, 1992. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1992 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "Using the Computer as a Reading Portfolio" by Judy Hetterscheidt, Lynn Pott, Kenneth Russell, and Jakke Tchang. From *Educational Leadership*, vol. 48, no. 6, p. 73, 1992. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1992 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "Assessing Reflective Responses in Journals" by Elaine Surbeck, Eunhye Park Han, and Joan E. Moyer. From *Educational Leadership*, vol. 48, no. 6, p. 25-27, 1991. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1991 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "What Teachers Learn from 'Kid Watching'" by Richard VanDeWeghe. From *Educational Leadership*, vol. 49, no. 7, p. 49-52, 1992. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1992 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "Teachers' Stories: Our Ways of Knowing" by Mary Renck Jalongo. From *Educational Leadership*, vol. 49, no. 7, p. 68-73, 1992. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1992 by ASCD. All rights reserved.

The Association for Supervision and Curriculum Development for "Learning From an Outcomes-Driven School District" by Tom Rusk Vickery From *Educational Leadership*, vol. 45, no. 2, p. 52-56, 1988. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1988 by ASCD. All rights reserved.

Phi Delta Kappan for "The Nation's Report Card Goes Home: Good News and Bad About Trends in Achievement" by Robert L. Linn and Stephen B. Dunbar. From *Phi Delta Kappan* vol. 72, no. 2, p. 127-133, October, 1990. Reprinted with permission of *Phi Delta Kappan.* Copyright 1990 by *Phi Delta Kappan.* All rights reserved.

Phi Delta Kappan for "Budgets, Politics, and Testing" by Chris Pipho. From *Phi Delta Kappan* vol. 72, no. 2, p. 102-103, October, 1990. Reprinted with permission of *Phi Delta Kappan.* Copyright 1990 by *Phi Delta Kappan.* All rights reserved.

Phi Delta Kappan for "Searching for Alternatives to Standardized Tests: Whys, Whats, and Whithers" by Walter Haney and George Madaus. From *Phi Delta Kappan* vol. 70, no. 9, May, 1989. Reprinted with

permission of *Phi Delta Kappan.* Copyright 1989 by *Phi Delta Kappan.* All rights reserved.

Phi Delta Kappan for "Assessment Literacy" by Richard J. Stiggins. From *Phi Delta Kappan* vol. 72, no. 7, March, 1991. Reprinted with permission of *Phi Delta Kappan.* Copyright 1991 by *Phi Delta Kappan.* All rights reserved.

Phi Delta Kappan for "But...the Curriculum" by Caroline Mamchur. From *Phi Delta Kappan* vol. 71, no. 8, April, 1990. Reprinted with permission of *Phi Delta Kappan.* Copyright 1990 by *Phi Delta Kappan.* All rights reserved.

Phi Delta Kappan for "Linking Restructuring to Authentic Student Achievement" by Fred M. Newmann. From *Phi Delta Kappan* vol. 73, no. 6, February, 1991. Reprinted with permission of *Phi Delta Kappan.* Copyright 1991 by *Phi Delta Kappan.* All rights reserved.

Index

A

Accountability testing, 44, 45-46
  safeguarding, 44-45
Accuracy, checking for, 216
Achievement. *See also* Authentic achievement
  trends in, 13-29
ACT (American College Test), 10, 13, 78
Active learning, 56-57
Alternatives to Standardized Testing, 90
America 2000, 9
American Association for the Advancement of Science—Project 2061, 22
American College Test (ACT), 10, 13, 78
American Council on the Teaching of Foreign Languages, 76
Assessment. *See also* Tests
  alternatives in, 59-67
  assessing, 9-11
  classroom, 41-44, 111, 112, 158-76
  constructed response, 162, 163
  direct-personal communication, 107
  expanding boundaries, 45-46
  methods of, 163-65
  need for better, 37-46
  performance, 107, 115-22, 162, 163, 164-65
  practicality of, 161
  process-focused, 162, 163
  product, 162, 163
  of reflective responses in journals, 181-86
  relationship between instruction and, 160
  safeguarding accountability, 44-45
  selected response, 162, 163
  state programs for, 31-36

Assessment literacy, 7
bottom line, 112-14
definition of, 102-3
differentiating levels of, 108-10
gauging, 103-6
programs for, 110-12
understanding, 106-8
Assessment of Performance Unit (APU), 115-16
Assessment planning chart, 167-69
sample, 174
using, 169-70
Authentic achievement, 136-39
depth understanding, 138
facilitating, 141-46
integration, 138-39
linking restructuring to, 133-46
prior knowledge base, 137-38
structures that facilitate, 141-46
access to tools and resources in, 143
collaboration, 142-43
discretion and ownership, 143
substantive conversation, 140-41
Authentic tests, 6-7, 76, 78-79
exhibitions of mastery, 71-74
performance-based, 74-76, 79
setting standards, 69-70
teaching to, 69-79

British Assessment of Performance Unit, 78
Budget, in California, 32-33

C

California Assessment Program (CAP), 32-33, 45
Cheating scandals, 11
Choices, in curriculum, 123-31
Civil Rights Act Title VII, 90
Classroom assessment, 41-44, 111, 112, 158-76
principles in, 158-60
process for planning and conducting, 165-67, 169-73,175-76
selecting appropriate approaches, 162-65
selecting methods, 160-61

steps for planning, administering, and interpreting, 166-67
Classroom-research-as-professional-development, 149-54
Collaboration, 142
Committee on Ability Testing of the National Research Council, 92
COMP, 78
Computers, in assessing reading progress, 187-89
Constructed response formats, 163
Contemplation, journals for, 184-85
Conversation, substantive, 140, 145
Council of Chief State School Officers, 141
Council of the Great City Schools, 36
Curriculum, choices in, 123-31

D

Debates, 172
Degrees of Reading Power (DRP) test, 94
Depth understanding, 138
Direct-personal communication, 107
Disciplined inquiry, 137
Discretion, 143
Douglas County School District's Higher Literacy Project, 149, 150-51

E

Education, U.S. Department of, 33
Educational achievement, judging, 9, 13
Educational Reform Act (English), 118
Education Commission of the States (ECS), 33, 88
Education Index, 88, 89
EEOC (Equal Employment Opportunity Commission), 90
Elaboration, journals for, 183-84
Elementary and Secondary Education Act (1965), 88
England, performance assessment in, 115-22
Equal Employment Opportunity Commission (EEOC), 90
Essay, 172
in standardized tests, 38
Everybody Counts, 22
Excellence, pursuing, 212
Exhibitions, 71-74
Experiences, recalling past, 217-18

General Certificate of Secondary Education (GCSE), 117
Grady Profile—Portfolio Assessment, 187, 188

H

Her Majesty's Inspectorate of Schools, 116
Higher-order thinking, 21, 27, 40, 43, 115, 203
Higher-Order Thinking Science Test, 76

Impulsivity, decreasing, 215
Instruction
  mastery learning-based, 206-7
  relationship between assessment and, 160
Instructional leadership, 208-10
Integration, 138-39
Iowa Tests of Basic Skills (ITBS), 26, 27, 161

Journals, assessing reflective responses in, 179-86

K

Kentucky, performance-based assessment in, 33-34
Kid watching, 149-54, 173
Knowledge
  application of, 157
  recalling past, 217-18

L

Language precision, 218-19
Leadership, instructional, 208-10
Learning
  active, 56-57
  application of, 6
  transference beyond, 218
Learning outcomes, 160, 201-3, 205
  instructional leadership, 208-10
  mastery learning-based instruction, 206-7
  overcoming historical obstacles, 211-12
  restructured responsibilities, 207-8
  staff development mastery model, 210-11
Literacy. *See also* Assessment literacy
  definition of, 53
  of thoughtfulness, 49-58

M

MacArthur, John D. and Catherine T., Foundation, 53
Majority students, and achievement gap, 19, 20, 21
Mastery learning-based instruction, 206-7
Mastery model for staff development, 210-11
Mathematics
  goals for achievement, 35
  results on NAEP, 21-25
Mathematics Report Card, 24
Matrix sampling, 32, 33, 45
Meta-cognition, 216
Minorities
  and achievement gap, 19, 20, 21
  and mathematics achievement, 25
  and test improvement, 26

N

Narratives
  as act of mind, 192
  as contribution to professional growth, 193, 195-98
  ethic of care in, 196-98
  as metaphor for change, 195-96
  power of, 191-92
National Assessment of Education Progress (NAEP), 9, 10, 13-17, 45, 115
  additional data, 25-29
  mathematics results, 21-25
  reading results, 17-21
  science results, 21-22, 25
  state-by-state comparisons in, 35
National Assessment Governing Board, 35
National Center for Fair and Open Testing (FairTest), 35
National curriculum assessment, 118-19, 121
National Defense Education Act (1958), 88
National Education Association (NEA), 90
National Governors' Association, 34-35
National Research Council, 22
National Writing Project, 76
Nation's Report Card. *See* National Assessment of Education Progress (NAEP)

O

Obstacles, overcoming historical, 211-12
Oral questions, 172
Outcomes. *See* Learning outcomes
Ownership, 143

P

Paper-and-pencil tests, 37, 107, 121
Pedagogical experimentation, 151
Performance-based assessments, 107, 163, 164
  diploma for, 79
  in Kentucky, 33-34
Personal Narrative. *See* Narratives
Politics, 33-35
Portfolios, 179-80
Practicality of assessment, 161
Preliminary Scholastic Aptitude Test (PSAT), 14-15
Prior knowledge base, 137-38
Problem posing, 216-17
Problem solving, 219-20
Process-focused assessments, 163
Product assessments, 163
Professional development, 153, 154
  classroom research in, 149-54
  contributions to, 193, 195-98
Project 2061 of the American Association for the Advancement of Science, 22
Project Spectrum, 50-51
  areas of cognitive ability examined in, 62
  classroom environment, 60-61
  disadvantages/advantages, 66-67
  early research results, 65-66
  features of, 61, 63-65
PSAT (Preliminary Scholastic Aptitude Test), 14-15
Public examinations, 116-18

R

Reading portfolio, using computer as, 187-89
Reading results, on NAEP, 17-21
Real-life situations, application of school-learned knowledge to, 218
Reflection journals, 185-86

Reliability, of assessment, 161
Responsibilities, restructuring of, 207-8
Restructuring
linking to achievement, 133-46
of responsibilities, 207-8

S

Scholastic Aptitude Test (SAT), 10, 13, 14-15, 93
School district, learning from outcomes driven, 205-12
School-learned knowledge, application to real-life situations, 218
Schools, creating thoughtful environments at, 58-59
Science
examples of standard assessment tasks, 119-20
goals for achievement, 35
results on NAEP, 21-22, 25
Science for All Americans, 22
Second International Mathematics Study (SIMS), 22
Selected response formats, 163
Selective abandonment, 6
Self-reflection, 151
Skills, application, 157
Solutions, persevering when not apparent, 214-15
Staff development mastery model, 210-11
Standard Assessment Tasks, 118, 119, 121
evaluation of, 120-21
Standardized tests, 5-6, 9-11, 90
alternatives to, 87-97
in classroom assessment, 42-44
complaints on, 89-90
content of, 38
essay format in, 38
examples of, 119
limitations of, 37-39
multiple choice format in, 38-39
narrowing of content, 38
negotiation of content, 38
teaching to, 39-41, 69-70
upsurge, 88-89
State assessment programs, 31-36
Student
choices of, concerning curriculum, 123-31

evaluating thinking abilities of, 213-20
involvement of, in learning process, 157-76
linking restructuring to achievement of, 133-46
thinking abilities of, 157
Substantive conversation, 140-41, 145

T

Teachers
effect of test-defined instruction on, 40-41
and kid watching, 149-54
narratives of, 191-98
professional growth of, 153-54
Teacher-made tests, 9
Teaching to the test, 39-41
Testing and Thoughtfulness (Brown), 53-58
Tests, 9-11, 31-33. *See also* Assessment
accountability, 44, 45-46
alternatives to, 83-97
authentic, 48-51, 69-80
course-specific, 74
paper-and-pencil, 37, 107, 121
results on, 13
standardized, 5-6, 9-11, 37-39
upsurge in, 88-89
Think-aloud activities, 173
Thinking
evaluating students' abilities in, 213-20
flexible, 215-16
Thoughtful communities, 150-51
Thoughtfulness, testing for, 53-58
Time scheduling, 144-45
Tools and resources, access to, 143
Transference, 218

W

Written list, 172
Written test, 171

SkyLight

Learn from Our Books *and* from Our Authors!

Bring Our Author/Trainers to Your District

At IRI/SkyLight, we have assembled a unique team of outstanding author/trainers with international reputations for quality work. Each has designed high-impact programs that translate powerful new research into successful learning strategies for every student. We design each program to fit your school's or district's special needs.

---

1 Training Programs

IRI/SkyLight's training programs extend the renewal process by helping educators move from content-centered to mind-centered classrooms. In our highly interactive workshops, participants learn foundational, research-based information and teaching strategies in an instructional area that they can immediately transfer to the classroom setting. With IRI/SkyLight's specially prepared materials, participants learn how to teach their students to learn for a lifetime.

---

2 Network for Systemic Change

ce

goal, Network consultants work with member schools to develop an annual tactical plan and then implement that plan at the classroom level.

---

3 Training of Trainers

BEST COPY AVAILABLE

IRI SkyLight
TRAINING AND PUBLISHING, INC.
2626 S. Clearbrook Dr., Arlington Heights, IL 60005
800-348-4474 • 847-290-6600 • FAX 847-290-6609

There are
one-story intellects,
two-story intellects, and three-story
intellects with skylights. All fact collectors, who
have no aim beyond their facts, are one-story men. Two-story men
compare, reason, generalize, using the labors of the fact collectors as
well as their own. Three-story men idealize, imagine,
predict—their best illumination comes from
above, through the skylight.

—Oliver Wendell
Holmes

SkyLight
TRAINING AND PUBLISHING, INC.

AUTHENTIC ASSESSMENT

A Collection

Edited by Kay Burke, a national presenter in the area of assessment, *Authentic Assessment: A Collection* brings you the best thinking on the topic. Featured authors include Rexford Brown, Arthur Costa, and Lorrie Shepard. The authors challenge us to ask, "What is authentic assessment? Is it just another 'buzzword' of the 1990s? What do we want our students to know twenty-five years from now? And how can we tell if they're learning that now?"

If we are to measure ongoing student growth and development, we need to shift from traditional paper-and-pencil, fill-in-the-blank, short-term memory Scantron tests to authentic assessment tasks that require students to perform skills like speaking, writing, performing, and problem solving. How do we strive for significant outcomes? How do we design criteria to measure those skills? Just open the pages of this book for thoughtful discussions about each of these questions.

IRI SkyLight
TRAINING AND PUBLISHING INC
2626 S. Clearbrook Dr. Arlington Heights, IL 60005

ISBN 0-932935-51-6
90000
9 780932 935519

BEST COPY AVAILABLE